

SEER Special Project #08
Development of High Resolution Population Distribution Data to Enhance Cancer
Prevention and Control Research
RFP No. NCI-PC-25014-20

Final Report
June 16, 2004

Iowa Cancer Registry/G. Rushton P.I.
The University of Iowa

Contains Final Report by subcontractor, Oak Ridge National Laboratory

I. Objectives of the Study

Currently, measures of cancer incidence for small geographic areas rely on population data for geographic areas defined by the U.S. Census. Consequently, these measures have traditionally only been available for Census areas or aggregations of them. This project developed age-sex population data for Iowa (Year 2000) for 90 meter square areas. Using geographic information science methods and Iowa SEER registry data, it computed and compared measures of colorectal cancer incidence using the population characteristics of Census areas with the same measures using the population characteristics of the 90 meter grid areas. The population data for the grid was provided by the Oak Ridge National Laboratory, Geographic Information Science and Technology group, as a subcontract to this contract.

Please note that the Appendices referenced in this report can be accessed as PDFs at the following website: www.uiowa.edu/~gishlth/UIORNL/

II. Progress to Date

A. Work Completed

We describe the results of analyses we performed that compare geographic measures of the cancer burden using the Census data and the 90 meter grid data. The final section summarizes our conclusions on the usefulness of this population data for these small geographic areas for types of geographic analyses that are typically done with such population data available only for Census tabulation areas.

All cancer data used in this study are from the Iowa (SEER) Cancer Registry. Each cancer incidence was geocoded using the common cancer registry practice of engaging a commercial geocoding firm: Geographic Data Technology. Their geocoding technique is to match the street address of the cancer case to its file of streets and roads for Iowa. This is an enhanced TIGER line file. Cases that could not be matched to the road network were located at the centroid of their five-digit zip-code.

Population data were taken from U.S. Census 2000 data files and, for the comparative purposes of this study, additionally from the LandScan USA data files.

1. Small Area Population Data

The Geographic Information Science & Technology Group (GIST) at the Oak Ridge National Laboratory made population totals for small areas in Iowa (90 meter square grid) that sum to their respective U.S. Census 2000 Block population totals. In brief, their method takes the Census Block level 2000 population data release and assigns these population counts to approximately 90 meter square areas within each Block on the basis of their scores on four basic data layers: proximity to roads, national land cover data, slope and weight for landmark polygon feature. The latter denote specialized land use

including parks, cemeteries, golf courses, and other recreational areas. Each grid cell receives a weight on each of these variables and the weights are combined in an empirically derived equation. “The total population of each block is allocated to each cell weighted by the calculated likelihood (population coefficient) of being populated.” From an overlay of their grid data and the boundaries of Census Blocks and Block Groups they computed population centroids for Census Blocks and Block Groups for Iowa. A detailed description of their method and the population data for Iowa that they provided can be found in their report which is attached as Appendix One and also available on the web site.

2. A Comparison of the Expected Number of Colorectal Cancer Incidences in Circular Areas with Populations Defined from Interpolated Block Group Data with Populations Defined with LandScan USA data

From the 90 health service area locations in Iowa, we selected twenty at random. Health service areas are defined by Wellmark, the dominant health insurer in Iowa, based on the hospital discharge data of its clients; see Figure 3 in Appendix 2. The scenario we examined was the difference between indirectly age-adjusted colorectal cancer rates computed for circular areas of 5, 10 and 15 miles respectively surrounding each of these 20 places. The numerator of these rates is the observed number of colorectal cancer incidences (between 1993 and 1997) in the three respective circles and is a constant in the two measures to be compared. We therefore focused on comparing the differences in the expected numbers of colorectal cancers in the three circular areas around each of the 20 places, since these differences reflected the differences between the estimated populations by age in the two denominator measures. For each of these areas we computed the expected number of colorectal cancer cases using the indirect, age-adjusted method. In the first case we used standard spatial interpolation techniques in which we estimated population totals in each five year age category from the proportion of each Census Block Group that fell within the circular area. In the second case we used population totals from the LandScan USA database. For this data we estimated the ages and sex distributions by applying to each 90 meter grid population total their respective proportions within their Census Block Group.

One key result (Fig.11 in Appendix 2, reproduced below) is that, over all of the 60 areas examined (20 places x 3 circular areas), the expected cases ranged within 10 percent of each other with many areas close to the 10 percent difference. There was a tendency for the difference to be greatest for the smaller places. The difference between the two population estimates is greatest for the circular area whose boundary cuts through the Census Block Groups that lie on the border of the built up area of the place. The larger the place, the further out from the center of the place this border lies.

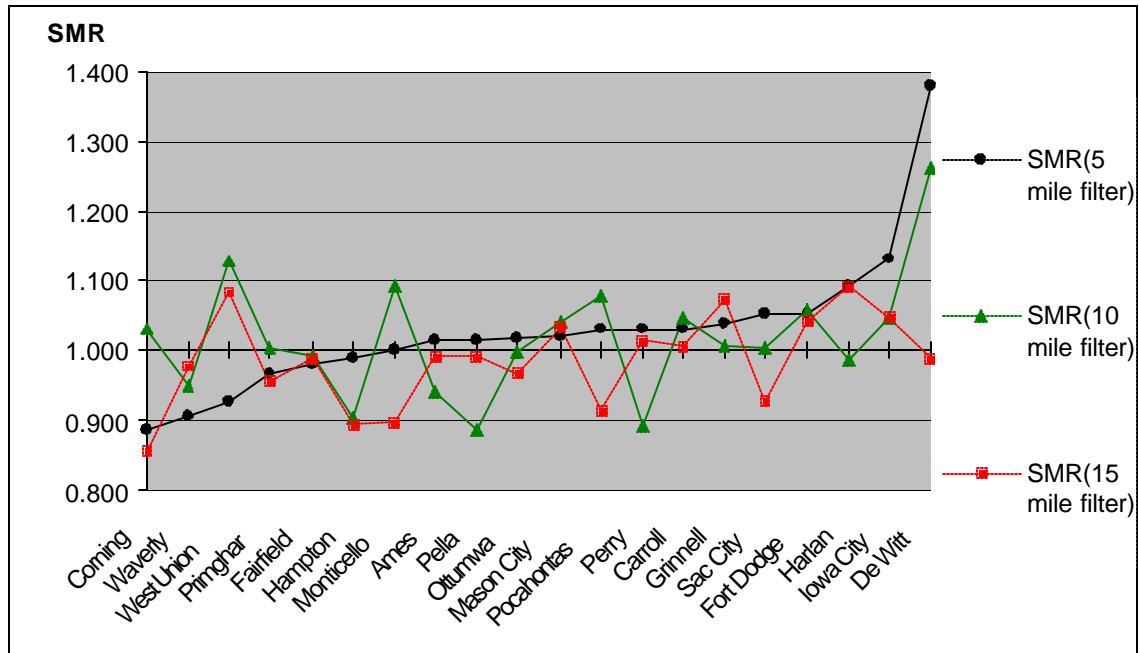


Fig. 1 (Fig. 11 in Appendix 2). The ratio of expected number of colorectal cancer cases using LandScan USA data / expected number of cases using Census data for 20 HSA areas in Iowa

A second key result is that for all filter sizes, the ratio of the expected number of cases from each of the two methods is furthest from unity when the expected number of cases is less than about 80. When the expected number of cases is larger than 100, the ratio of the expected numbers of cases from the two methods is close to unity.

A third result, which can be seen in Fig. 2, is that the ratio (expected LandScan / expected Census) is more frequently greater than one, indicating that population estimates from LandScan for areas such as those examined in this experiment are generally larger for LandScan USA data sources than when estimated from Census sources. We emphasize that this conclusion is for areas such as examined here, noting that all twenty areas selected focus on a town or city. From towns and cities, densities of population generally decline in an exponential form with respect to distance from center of town. Thus, Census Blocks that are found in the border area of the spatial filter will tend to have larger population densities on the side of the place than on the edge furthest from the place. The spatial interpolation assumption is that population density is uniform within the Census Block Group area when in fact densities are higher on the inner city side than the outer city side of the Census area.

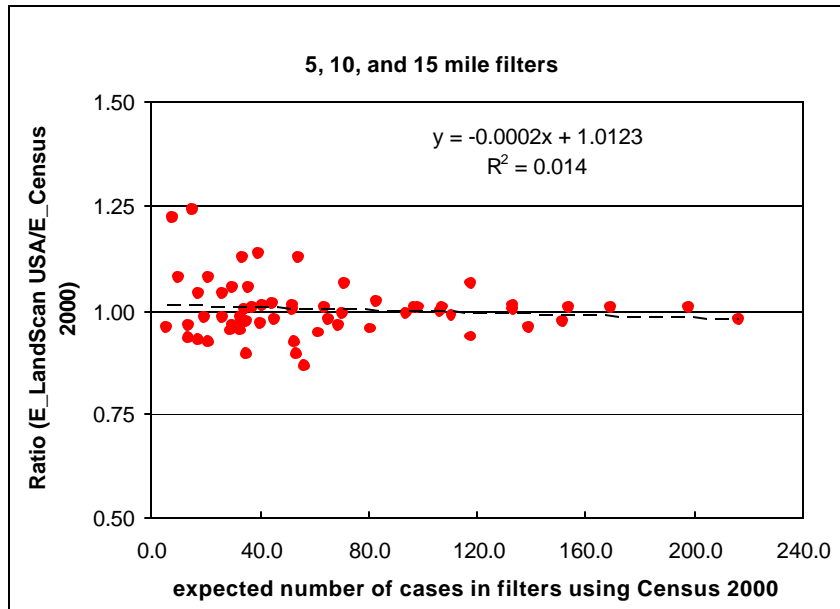


Fig. 2 (See Fig. 16d, Appendix 2). Cancer rates are most different when the expected number of cases is less than about 100.

From these experiments we conclude that there can be considerable differences in the population characteristics of areas that do not conform with Census designated areas between estimates computed through aggregations of LandScan USA blocks and estimates computed using the areal weighting assumption method commonly used for this purpose in the GIS literature. Since these circles represent the sizes of areas for which cancer rates are typically reported, we conclude that these differences do represent a serious source of bias in cancer rates that to a large degree could be eliminated by using LandScan USA type of small-area population data. The methodology we used to produce the results for this section and more details of the results can be found in Appendix two on the website.

3. Estimating Distances of People to Facilities

One of the expected advantages of using small area population data is that when the purpose is to estimate the distances of people to facilities, distance estimates made from the centroids of small areas will be more accurate than distance estimates made from the centroids of larger areas. If for example, distances from the centroids of Census Blocks, weighted by their respective populations, to defined facilities such as breast screening facilities, are computed, how much less accurate are they from distances estimated from the centroids of LandScan 90 meter grids? There is a large theoretical and empirical literature on this subject. The theoretical literature began with Hillsman and Rhoda (1978) and Current and Schilling (1987). These works showed that three kinds of errors are introduced by the representation of continuously distributed people by discrete areas. The three types of errors are concisely described in Current and Schilling (1987, p.98).

Two of these errors result in distances being underestimated and the third error overestimates distances. Therefore a general conclusion is not possible that in all cases the use of centroids from which to estimate distances results in a bias to overestimate or underestimate the true distances. So, what is known? We know from Francis et al. (1999, p.83) that “there is a nonlinear relationship between aggregation error and the number of aggregation points. After a certain point, increasing the number of aggregation entities gives very little payoff in terms of a decrease in aggregation error.” Their work further showed that the ratio of q/p is critical for determining the expected severity of distance error. In this ratio q is the number of spatially aggregated units (in this case, Census Blocks, Block Groups or Zipcodes) and p is the number of facilities. The more locations have facilities and the fewer locations are used to represent demand, the greater the distance error. The Francis et al. article shows that when q/p is less than 20, the percent relative distance error is around five percent. The literature also shows that spatial population medians are more appropriate than spatial population centroids in representing spatially aggregated population data for the purpose of making population weighted distance estimates to facilities. Population centroids minimize the sum of squared distances of population to themselves, whereas population medians minimize the sum of distances of population to themselves. Since geographical accessibility is generally measured as population weighted distances, spatial medians provide more accuracy than population centroids. Algorithms for computing spatial medians for weighted points in the plane—such as LandScan USA data, are well-known, see Kuhn and Kuenne 1962 and Weisfeld 1937.

There is also a large empirical literature on this subject. Error in estimating distances from spatially aggregated data is often known in this literature as “aggregation error.” As Hewko et al (2002, p.1185) noted, “aggregation error arises when, for the purpose of distance calculations, a single point is used to represent a neighbourhood, which in turn represents an aggregation of spatially distributed individuals.” They measured “neighborhood spatial accessibility” (NSA) in a Canadian city and concluded (p.1185) that “aggregation error is particularly problematic when measuring NSA to amenities that are abundant and have highly localized service areas, such as playgrounds. We recommend that, when analyzing NSA to these types of amenities, researchers integrate finer resolution data to indicate the spatial distribution of individuals within neighborhoods better, and hence reduce aggregation error.” Such highly localized service areas will have ratios of q/p less than 20 whenever Census areas are used as the basic spatial units. It is for such situations that fine, spatial resolution population data, such as the LandScan USA grid data, will be most useful. Cancer control and prevention programs often focus on the geographical accessibility of populations to local facilities. Such applications would benefit from the fine geographic detail of LandScan USA population data.

4. Errors in Spatial Coverage when Zip Code Areas (ZCTAs) are Assigned to Spatial Filters Based on Their Centroids—implications for cancer maps

This part of our research investigated the spatial coverage errors that occur when circular areas are used to aggregate zonal data using the principle that all zones whose centroids lie within the circle are aggregated. Such errors are avoided by using LandScan USA population data. We computed these errors for 2239 circular areas for each of five different distance radii located at grid points on a five mile grid covering all of the State of Iowa in which these circles were fully covered within the state. Appendix Four (accessible on the website) describes the results. Two kinds of errors were measured: the error of not counting population within the circle when zones whose centroids lie outside the circle have part of their area within the circle (inside-error) and the error of counting population outside the circle when the population of zones whose centroids are within the circle have part of their areas outside the circle (outside-error). We found that an average of 25 percent of the area of the ten mile filter circles lies either inside the circles but not counted or outside the circles and counted. The error of omission or commission changed rapidly across the map. No discernible spatial pattern of these errors was found. The significance of these results are that these coverage errors can be removed by using LandScan USA population data.

B. Work not completed that was proposed

All but two of the analyses contracted to be produced in this report have been completed—see “Addendum 29, Statement of Work—Attachment 1.” The not completed analyses consist of #3 in the statement of work: “test the hypothesis that a person’s cancer stage at first diagnosis is related to the demographic characteristics of their local area as defined below,” and the proposed analyses to compare the effects of replacing geometric centroids with population centroids when estimating distances of people to facilities. Section #3—see page 5 above, substitutes for analyses using ORNL’s population centroids by describing the considerable theoretical and empirical work that exists on this subject. It is clear from the literature described in section three that future work to reduce errors in estimated distances from aggregated spatial data should use spatial population medians, not population centroids, for purposes of more correctly estimating distances.

The LandScan USA data had some inherent limitations. These problems are described below. The effects of U.S. Census disclosure limitation methods, in particular, had not been anticipated. Although these methods do not affect Census Block population totals, they do affect Census Block age and sex totals such that we believe these should not be used as controls for determining age-sex totals in LandScan USA data. Consequently the University of Iowa group assumed the task of using the Census Block Group age-sex proportions to interpolate age-sex numbers to the LandScan USA 90 meter grid areas.

III. Problems Encountered and Resolution of the Problems

We encountered a number of problems in using the LandScan USA data. Because the ORNL GIS Group told us that technical information of their approach was proprietary, we found it necessary to evaluate its accuracy in a sample area. We made a detailed

analysis of the characteristics of their population grid in Carroll County, Iowa, to determine its usefulness for the purposes of this project. This rural County, located approximately 70 miles northwest of Des Moines, Iowa, is a county for which we have orthophoto imagery taken in the spring of 2001. The spatial resolution of this imagery is at the scale of two feet per pixel and the GIS Coordinator for the County has encoded all rural addresses, roads and the boundaries of all Census areas. At this geographic scale, farm buildings and residential buildings are clearly visible.

Our first analysis of the LandScan population grid for this County showed us that the 90 meter grid of population often did not locate population at the end of the short, private roads that typically link from the public road system to rural residences. After discussing this finding with the ORNL GIS Group staff, their review, they told us, showed that a TIGER Line File code that indicates a road junction with a private road, was encoded incorrectly by the U.S. Census for Carroll County and for many other Iowa Counties. ORNL then later provided us with a new data file in which this error had been corrected and their analyses re-run. This evaluation of LandScan USA population data based on our human interpretation of orthophotographs from one county in Iowa led us to conclude that the quality of this population grid is quite satisfactory. This interaction with ORNL GIS Group staff showed us that the LandScan USA method included additional factors to those included in their earlier statement of methodology. Consequently, we must advise that the description of their methodology, Part II of this report, is less than complete as a scientific description of their methodology. We note below three problems with their data that affect its appropriateness for research in spatial epidemiology. It should be noted that the first problem relates directly to LandScan USA methodology, the second problem relates to lack of quality in Census TIGER data, and the third problem relates to Census disclosure limitation and consequent impacts on the accuracy of the age-sex estimates in the population grid data and Census Block data.

The Population Grid's Congruence with Census Block Boundaries: the need to spatially interpolate between two different polygon descriptions.

When the LandScan USA grid is placed over Census Block boundaries the boundaries of grid cells do not follow Block boundaries. This is unavoidable, of course, since the grid cells are defined in fractions of minutes of latitude and longitude and the Census Blocks are polygons of arbitrary size and shape. Figure 1 in Appendix 5 illustrates this misalignment between the two kinds of areas. The ORNL method description does not explain how this misalignment is resolved. As Appendix 1 shows it appears that the ORNL method uses the proportion of the grid cell that lies in the Census Block as a weight that brings in population from the two or more Blocks with which it is associated. Some type of spatial interpolation function is being used though precisely which is not clear.

The Population Grid Registration: the problem of misaligned Census Block boundaries.

By overlaying the LandScan USA population grid directly over the orthophoto map of Carroll County and also superimposing the Census TIGER Linefile boundaries of Census Blocks, we were able to see that the literal assignment of the grid cells to their respective Census Blocks were subject to considerable spatial misalignment. In this sample area, the congruence of the true Census Block boundaries and the TIGER Census Block boundaries varied from very close to inaccuracies of approximately twenty percent in the worst cases. That is, because of spatial errors in the boundary lines of Census Block areas in the TIGER linefile, error that is not known (but less than 20 percent) exists in the assignment of the area observed in the source imagery for the LandScan USA population grid to its true Census Block counterpart. We note that this error does not affect the accuracy of Census tabulations since these are based on the topological, not the geometrical, characteristics of the Census area boundaries. That is, Census tabulations for areas are based on a look-up table in which the unique census tabulation areas of each street segment in the TIGER Line File is coded. The LandScan USA grid and its interpretation of the U.S. Census areas do rely on the geometrical accuracy of Census area boundaries which, unfortunately, are sometimes poor. In Appendix five, we illustrate this problem—see Figure 2 there. The misalignment of the TIGER block boundaries can be resolved in future LandScan USA products by using commercially available data, such as GDT Dynamap or Navtech which have higher positional accuracy or by using publicly produced enhanced TIGER Line files generally based on orthophoto maps similar to the one we used in this case of Carroll County, Iowa. Although the quality of the LandScan USA data will be significantly enhanced by these alternative data sources, in many cases the cost of these data sets (in thousands of dollars) can significantly impact the scope and feasibility of any project of this scale. A recently published NRC report discusses U.S. Census plans for Census 2010 and the improved TIGER Line files that both it and the American Community Survey will require—see National Research Council, 2003, Chapter 3.

The Population Grid: Effects of Census Disclosure Limitation Methods

Data swapping was used to protect the confidentiality of the Census 2000 tabulations (U.S. Census Bureau, 2003; Steel and Zayatz 2003, Zayatz 2002). The Census Bureau considers the disclosure limitation methods it used in the 2000 Census to be proprietary information and the public has access only to the report cited above which is an abridged version of a final report which is available to selected Census Bureau personnel on a need-to-know basis. The abridged report contains some information which permits a partial evaluation of the effect of data swapping on the accuracy of the population grid data. Some characteristics of the swapping procedure were described in the above report.

“The procedure was performed on the underlying microdata, and all tabulations from the 100 percent (short form) and from the sample (long form) data were created from the swapped files. It affected pairs of households (or partnered households) where one or both of those households had a high risk of disclosure. The set of census households that were deemed as having a disclosure risk was selected from the internal census data files. These households were unique in their geographic area (block for 100 percent data and block group for sample

data) based on certain characteristics. The data from these households were swapped with data from partnered households that had identical characteristics on a certain set of key variables but were from different geographic locations. Which households were swapped is not public information.” (p.v).

As their paper explains, the Census method used restricted swaps. Restricted swaps are swaps between households that are identical on a number of characteristics.

Unfortunately, the Census does not reveal which characteristics between swapped households are forced to be identical in each pair to be swapped. However, it does make clear that household size is one such restriction on all swapped pairs.

“Unrestricted data swapping could have the effect of smoothing the data, reducing its natural concentration and decreasing the empty parts of the tables. We found no such effect for our procedure, primarily due to the forced agreement of household size.”(p.7).

I understand this statement to mean that the number of people in each Census Block or any other census area entity will be unaffected by the Census disclosure limitation method but that any other attribute of these persons may well be affected by disclosure limitation. From the viewpoint of this project, this means that the sex and age characteristics of people in Blocks and even Block Groups may be inaccurate because of disclosure limitation. Laura Zayatz, co-author of the Census report on effects of disclosure limitation procedures on Census 2000 tabular data products, in personal correspondence with Qiang Cai, research assistant to Dr. Rushton, added that the numbers of persons below 18 and the numbers of persons above 18 are also correct at all levels of Census geography. One can infer from this that, in addition to ensuring that the number of persons in a household is the same in all swapped records, the procedure also ensures that the number of persons below 18 and the number above 18 in a household is also the same in all swapped pairs of Census records. The Census Bureau will not provide a list of other attributes of households, if any, that are matched for swapping purposes nor will they provide information on the geographic distance between the swapped records. They do not even guarantee that the swapped records will be within the same County, though Laura Zayatz did note that “Most of the swapping would be done within county, but there is no guarantee.” Thus, the disclosure limitation procedure will affect the accuracy of the age-sex estimates in the population grid data, as discussed below.

Making Age-Sex Estimates from Household totals: the ORNL data does not contain age-sex information

Since the inference procedure used by ORNL did not assign age or sex to the 90 meter square areas, we estimated these totals using two different methods. The first is a commonly used estimator consisting of allocations of age-sex characteristics to individual areas on the basis of their proportions in the Census Block Group as a whole. We used Census Block Group proportions rather than Census Block proportions because the proportions in five-year age/sex groups are subject to inaccuracies introduced into the Block level proportions by the Census disclosure limitation procedure described above. The Census Block Group proportions, of course, also could contain error from the same source, but this is much less likely to occur with these areas.

The second method used is derived from Tobler's pchnophylactic method. The high resolution population grid from LandScan USA was very useful in implementing this second method as it provides very good estimates of total population in the form of an approximately continuous surface. We believe this method holds much promise for the future. It is described in detail in Appendix Three. The results we reported in the section above are based on the first of these two methods.

IV. Conclusions on LandScan USA methodology for the purposes of this project.

In the above sections we have noted three problem areas that each contribute to probable error in the process of spatially disaggregating Census Block populations to the fine geographic scale of a 90 meter grid. The extent of the resulting error is not known though clearly it will depend on a number of characteristics of the area for which the disaggregation is made. The more diverse the Census Block areas, the more disaggregation of the population will improve detailed geographic measures of cancer rates.

A comparison of LandScan USA and US Census 2000 data for computing the expected numbers of colorectal cancer cases from 1993 to 1998 in different radius filters around 20 health service centers in Iowa using the two data sets offers, among others, the following two conclusions:

1. The difference of population estimates is the primary source of difference in the expected number of cancer cases calculated using different population data. The highly heterogeneous population distribution in space means that population estimates for small areas can vary geographically. In our example, absolute difference of total population estimates in a five mile filter area can vary from less than 50 to over 6,000 people. Clearly, the high level of geographic specificity in the LandScan USA data is helpful in giving a high level of accuracy to the estimates of expected numbers of cancer cases in these custom-defined areas.
2. The differences in expected number of cancer cases in the 20 HSA areas varied between 10% less to 40% more for circular analysis extents of 5, 10, and 15 miles in radius. Since these circles represent the sizes of areas for which cancer rates are typically reported, we conclude that these differences do represent a serious source of bias in cancer rates computed from misaligned population that, to a large degree, could be avoided by using small area population data such as LandScan USA.

The current limitations of the LandScan USA methodology, however, need to be addressed. In addition to quantifying the improvements resulting from better quality input data, the subjective nature of the (empirical) weighing methodology needs to be statistically refined. Consequently, understanding the errors and spatial uncertainties in the population distribution data should be methodically addressed. Investments in future research focusing on these important science issues, which go beyond the utility of developing a database, will significantly improve our scientific understanding of the

sensitivity of such population disaggregation approaches and will provide better confidence in such data sets as they are applied to spatial epidemiology.

In reference to aggregation error discussed above, two key questions regarding the usefulness of LandScan USA data for analyzing geographic patterns of cancer rates are:

1. How frequently will local accessibility of people with respect to cancer treatment or screening facilities be important?
2. How frequently will measuring the distance of people to possible sites of contamination that might increase population cancer rates be important?

Should the frequency and importance of these two questions be significant, it is evident that small area population such as LandScan USA could potentially offer critical value in answering these questions.

V. References

- Current, J.R. and D.A. Schilling. 1987. "Elimination of source A and B errors in p-median location problems." *Geographical Analysis* 19:95-110.
- Francis, R.L., T.J. Lowe, G. Rushton and M.R. Raco. 1999. "A synthesis of aggregation methods for multifacility location problems: strategies for containing error." *Geographical Analysis* 31 (1) 67-87.
- Geertman, S.C.M., V. E. J.R. Ritsema. 1995. "GIS and models of accessibility potential: an application in planning." *International Journal of Geographical Information Systems* 9:67-80.
- Hewko, J., K.E. Smoyer-Tomic, M.J. Hodgson. 2002. "Measuring neighbourhood spatial accessibility to urban amenities: does aggregation error matter?" *Environment and Planning A* 34:1185-1206.
- Hillsman, E.L. and R. Rhoda. 1978. "Errors in measuring distances from populations to service centres." *Annals of Regional Science* 12:74-88.
- Kuhn, H. and R. E. Kuenne. 1962. "An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics." *Journal of Regional Science* 4:21-34.
- Leung, Y., and J. P. Yan. 1998. "A locational error model for spatial features." *Int. J. of Geographical Information Science* 12:607-20.

- National Research Council. 2003. Chapter 3, Modernizing Geographic Resources, *in Re-engineering the 2010 Census: Risks and Challenges*, Washington D.C., NRC Press.
- O'Kelly, M.E., and Horner, M.W. 2003. "Aggregate accessibility to population at the county level: U.S. 1940-2000." *Journal of Geographical Systems* 5:5-23.
- Shi, W.Z. 1998. "A generic statistical approach for modeling errors of geometric features in GIS." *Int. J. of Geographical Information Science* 12:131-43.
- Talen, E. and L. Anselin. 1998. "Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds." *Environment and Planning A* 30:595-613.
- Steel, P. and L. Zayatz 2003. The effects of the disclosure limitation procedure on Census 2000 tabular data products (Abridged). Washington, D.C., U.S. Census Bureau, April 15, 2003.
- Weiszfeld, E. 1937. "Sur le point pour lequel la somme des distances de n points donnees est minimum." *Tohoku Mathematical Journal* 43:355-386.
- Wenzhong, S., C.K.Cheung and C. Zhu, 2003. Modelling error propagation in vector-based buffer analysis. *Int. J. Geographical Information Science* 17:251-71.
- Zayatz, L. 2002. SDC in the 2000 U.S. Decennial Census. In J. Domingo-Ferrer (Ed) *Inference Control in Statistical Databases: From Theory to Practice*. Heidelberg, Germany, Springer-Verlag, LNCS, Vol. 2316, p. 183-202.

These Appendices are PDFs on the website www.uiowa.edu/~gishlth/UIORN/

- Appendix 1 Development of High Resolution Population Distribution Data to Enhance Cancer Prevention and Control Research, by Budhendra Bhaduri, Edward Bright and Philip Coleman, Geographic Information Science and Technology Group, Oak Ridge National Laboratory. (12pp.)
- Appendix 2 A Comparison of the Expected Number of Colo rectal Cancer Incidences in Circular Areas with Populations defined from Interpolated Block Group Data with Populations defined with LandScan USA data, by Gerard Rushton, Qiang Cai and Michele West. (21pp.)
- Appendix 3 A Methodology for Estimating small-area populations by age and sex based on the Integration of Spatial Interpolation and Statistical Inference by Qiang Cai.(47pp.)
- Appendix 4 Errors in Spatial Coverage when Zip Code Areas (ZCTAs) are Assigned to Spatial Filters Based on Their Centroids—implications for cancer maps, by Gerard Rushton and Qiang Cai. (14pp.)
- Appendix 5 The Congruence of LandScan USA Boundaries with U.S. Census Block Boundaries, by Gerard Rushton and Qiang Cai. (4pp.)