Paper Title:
Effect of Altitude and Smoking on Low Birth Weight
Devon Williford
GIS Developer/Analyst
Colorado Department of Public Health and Environment
(303)692-2247
devon.williford@state.co.us

Paper Abstract

The State of Colorado is studying the effects of altitude and cigarette smoking on the occurrence of low birth weight, to identify specific communities of high risk and concern. The incidence of low birth weight in Colorado has been higher than the national average for each year since 1950, which constitutes a significant health issue. Using confidential information recorded through state birth certificates, the Colorado Department of Public Health and Environment (CDPHE) is examining the spatial distribution of more than 560,000 birth records, containing the address and cigarette smoking habits of each mother, from 1993-2002. These records were geo-referenced, given an elevation attribute, and analyzed in a GIS environment, utilizing statistical tools for detecting significant clusters of low birth weight. The results of the analysis allows medical professionals to view the occurrence of low birth weight in a more quantitative environment than has previously been available.

Introduction: Creating a Detailed Picture

Smoking habits of pregnant mothers have always played a factor in the incidence of low birth weights (<2500 grams or 5 lbs 9 oz). In Colorado, high altitude and hypoxia have also been identified as contributing to low birth weights. Recent studies have always been available in table format, producing statistics of low birth weights by census tract or county. Until recently, few attempts have been made to display and analyze these characteristics spatially.

In order to create a detailed picture of the effect of smoking and altitude on low birth weights, the Colorado Department of Public Health and Environment (CDPHE) is working with the University of Colorado Denver and Health Sciences Center, and the Colorado Center for Altitude Medicine and Physiology (CCAMP), by integrating the use of Geographic Information Systems (GIS) software and analysis with birth certificate records that are collected, secured, and maintained by CDPHE. 560, 840 birth certificates were geo-referenced by address of the mother, which includes all the singleton births (low, normal, and high birth weight) with certificates from 1992-2003 in Colorado. By geo-referencing these records, we were able to analyze other characteristics about pregnant mothers such as smoking habits, altitude of their address, and the numerator for this study, low birth weights. We were also able to remove the influence of elevation, by looking at low birth weights within specific elevation groups. This was helpful in searching for other common characteristics of this problem.

Using *Centrus* address modeling, *SaTScan* statistical software and *ArcGIS 9*, we were able to analyze the spatial distribution of low birth weight, detect spatial clusters, and evaluate the results to see if the clusters were statistically significant. The *ArcGIS* software was also used to create maps of these clusters that linked altitude characteristics, and health services data together, and produced enhanced graphic representation of results as compared with previous analyses.

Develop a systematic approach for deriving good location data and assigning elevation:

Birth certificates provide information on maternal smoking and other risk factors that may contribute to low birth weights, but do not initially contain altitude information. By geo-referencing these records to an exact location, we could examine, or track, the incidence of low birth weight over space and time, as well as examine other environmental factors that are spatially enabled. The process of creating spatial data from the birth certificate records began by developing a theoretical approach to geo-referencing, or locating, the address of the mother on each birth record. Initially, it would seem impossible to geo-reference all 560,000 birth certificate records to their exact location so we decided to organize the process into a two-phased approach. Standardization of field names for geo-referencing data that contains address information, and the documentation of the task of assigning spatial data, in this case both the accuracy and processes used to locate each birth certificate record address, are critical parts to this analysis. When this geo-referencing process is repeated for analysis of other public health datasets, it will save large amounts of time and resources, and ultimately provide a better tool for prevention and epidemiological research.

The first phase was to process the data through the *Centrus* software address coding module. This would be the first pass at geo-referencing the records. The Address Coding module compares the addresses from the birth certificate records to records in the United States Postal Service (USPS) ZIP+4 Directory and an enhanced street network file (GDT data). If there is a match between the birth certificate record address and the street network in *Centrus*, a very accurate latitude and longitude coordinate (NAD 83) for the birth certificate record is returned. In Phase I, *Centrus* returned housetop locations (latitude and longitude coordinates from the street network) to all but 28,276 records or 4.74% of our initial dataset.

At first glance it appeared that the majority of these records that did not match were located in rural or mountainous areas of the state. Figure 1 below shows the distribution of urban areas across the state on top of an image showing darker shades that represent more relief.

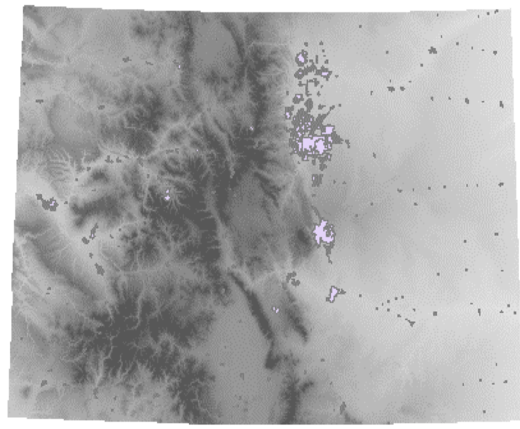

Figure 1. Map showing cities and towns in Colorado superimposed on an elevation model of the state

These records also contained some incomplete information when compared to the records that did match including insufficient address information, street designations that did not conform to the street network that *Centrus* utilizes, and trailer park names and/or apartment complex names instead of street addresses. When examining the 28,276 records that did not geo-reference in Phase I more closely in a statistical analysis, it was noted that the risk factors for low birth weight, including low maternal age, education, and smoking were significantly more prevalent than in the birth certificate records that did successfully geo-reference.

The records that did not match during Phase I provided opportunities to recognize, document, and edit the birth certificate records, now in a GIS database format. Maps, gazetteers, and use of the address standardization features (Zip Code centroid, etc.) in *Centrus* were utilized in Phase II of the geo-referencing process. In some instances, a single address was used to identify birth certificate records identified at trailer parks or apartment complexes. This process retrieved locations that were close to housetop locations in their accuracy. By the end of Phase II, around 35% of the 28,276 birth certificate addresses that were unaccounted for at the end of the Phase I were resolved.

The next phase in obtaining elevation for each record was to use *ArcGIS 9* software with the *3D Analyst* extension. Birth certificate records that did have an assigned coordinate were overlaid onto the 30M Digital Elevation Model (DEM) for Colorado. The z-coordinate or elevation was added to the database and calculated for each record. Based on standard errors in using DEM's with point data, this elevation value is usually accurate to within 40 meters of a point's true location on the ground. Based on the lack of relief (5% slope or less) for a particular town or zip code, some birth certificate addresses that did not obtain coordinates directly from *Centrus* in Phase I or Phase II were directly assigned elevations from the DEM to their relative locations.

After all of the records were assigned an elevation, it was elected to also assign each record an elevation category: 3000-5000ft (1), 5000-7000ft (2), 7000-9000ft (3), and above 9000 ft (4). This was done so that during our spatial analysis of the data we could not only look at the distribution of low birth weights across the state, but also within specific elevation groups that contain similar elevations. This would allow us to treat elevation as a continuous variable and look for patterns that may not be altitude related. The elevation categories we chose to use were defined by previous research done on low birth weights in Colorado by Lorna Moore in *Small Babies Among Big Mountains*. (2001)

Using *SaTScan*'s Bernoulli Probability Model

When the process of geo-referencing the birth certificate records was completed, the total number of singleton birth mothers that had been located was 560,183. We also were able to use 557,161 of these records for examining the smoking habits of the mother. The next process was to take this spatial data and analyze it to detect significant clusters, or groups, of low birth weights across time and the geographic boundary of Colorado, using *SaTSCan* software. *SaTSCan* software is available for download from the internet (http://www.satscan.org) for no charge, and is used frequently for spatial analysis of public health data.

SaTSCan allows the user to take advantage of two probability models that can identify statistically significant clusters of incidence, the Bernoulli and Poisson. For this analysis, we utilized the Bernoulli Model. The Bernoulli Model requires a dataset to have cases (low birth weights) and controls (all births). Area-based information such as the location of mothers during their pregnancy may be aggregated and represented by one single geographical location, in this case a 7.5-minute grid of the state of Colorado. Comparing the characteristics of each cell was a more manageable solution than processing 560,183 records through the Bernoulli model. Using this methodology, we developed four case files and two control files for our initial study as listed below. This would provide a detailed picture of low birth weights, very low birth weights (<1500 grams), smoking, and how smoking and low birth weights interact.

	Count
Original Birth Certificate Records 1992-2003	597204
Births Not Used After Geo-Referencing	36993
Births Falling Outside 1:24K Grid	28
Control File I: All Singleton Geo-Referenced Births Inside Grid	560183
Control File II: Control File I With Smoking Yes/No Attribute	557161
Case File I: Low Birth Weights and Very Low Birth Weights	38461
Case File II: Very Low Birth Weights	5033
Case File III: Smoking Mothers	59623
Case File IV: Smoking Mothers with Low Birth Weights	7581

Figure 2. Table Showing Number of Records Used in *SaTScan* Analysis

In order to create our case and control files to run through the Bernoulli model in *SaTScan* we used the "join data from another layer based on its spatial location" functionality in *ArcGIS 9*. In order to build the case file, 38,461 low birth weight locations that were geo-referenced were overlaid onto the 1,792 grid cells (1:24,000 USGS DRG index grid) for Colorado. For building the control file, 560,438 point locations were joined onto the 1,792 grid cells. Each grid cell was given the mathematical average of characteristics of all births that were inside the grid cell. These characteristics included a count of the births falling in each grid cell, an average latitude and longitude of the births in each cell, an average elevation category, an average birth weight, and an average smoking index. The 1,192 grid cells that contained one or more birth weights (low or other) are shown in green below (left). A demonstration of how the average latitude and longitude coordinate of all the births within a grid cell is shown at below (right):

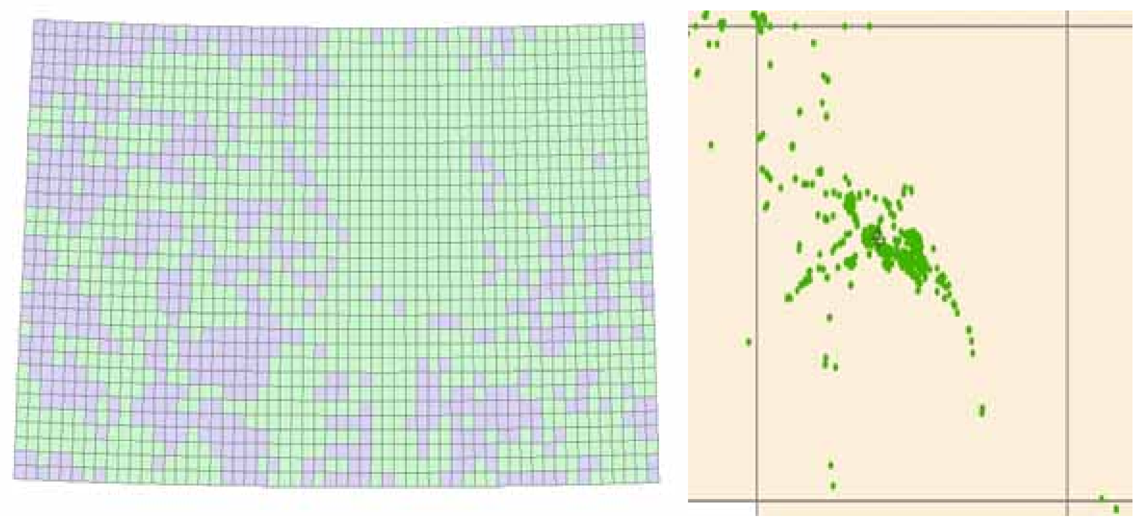

Figure 3. 7.5 minute grid of Colorado with cells in green that contain births (left), and lat/long centroid location for all births in an individual cell (right).

A coordinates file is also required for running the Bernoulli model. The coordinates file provides the geographic coordinates for each low birth weight. In our analysis, the coordinates file would also be the control file (all births).

After designating the case, control, and coordinate files in the *SaTScan* interface, the Bernoulli model scanned our case grid versus the control grid to look for areas with high rates (clusters) of low birth weights. We chose to look for areas of high rates of low birth weights, within similar elevation categories. *SaTScan* imposes a circular "window" on the surface of our grid and turns the window on each of several possible grid points positioned throughout the surface. The radius of the window varies continuously in size from zero to an upper limit exceeding the surface. The circular window is flexible both in location and size. This creates a large number of distinct geographical circles with different sets of neighboring data locations within them. This is essentially how *SaTScan* looks for clusters of incident data or expects to see clusters of incident data. *SaTScan* outputs a data base file (.dbf) containing the grid cell id of cells that were statistically significant that we could join that table into our grid layer. We used a p value of .05 or less in determining which clusters were statistically significant or not.

The following *SaTScan* settings were selected before running the software program:

Analysis: Purely Spatial
Probability Model: Bernoulli
Scan For Areas With: High Rates
Output: Cluster Location, Location Information, Risk Estimates

SaTScan produced two important tables (.dbf files) that were joined with the GIS grid cell layer. The grid cell layer now contained the cluster identification, p-values for each cluster denoting it's significance, and values for observed and expected cases for each cluster. This information could then be displayed on maps.

ArcGIS 9 Map Output

In *ARCGIS 9*, we joined the *SaTScan* cluster output (.dbf files) to out initial 1:24000 control grid in order to show where the significant clusters were. We also overlaid this information on a county boundary layer to add a geographical reference to the map. We decided to add cells that contained births to the map with a gray color to show where in the state we did or did not have data for this study.

The map below shows the significant low birth weight clusters that were defined by the *SaTScan* software. Three significant clusters of low birth weight are identified. The most highly significant cluster occurs in the metropolitan Denver area, with two other clusters along the northern and south-central mountain ranges. In the most significant cluster, *SaTSCan* expected to see 10,760 low birth weights within the cluster, but it actually observed 12,251 cases. It is important to understand that this is looking for clusters of low weight births within similar elevation categories.

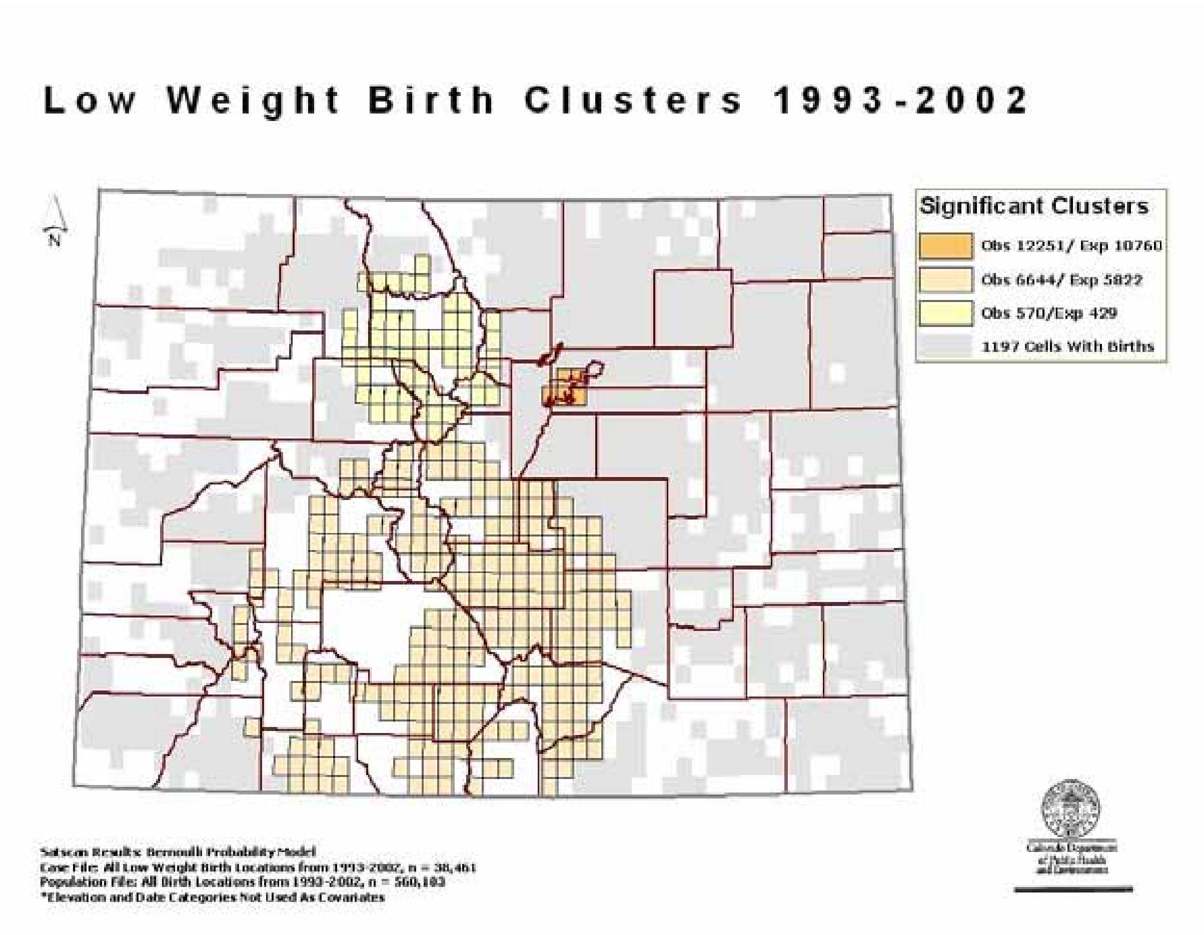

Figure 4. Low Weight Birth (<2500 grams) Clusters in Colorado (1993-2002)

The next map below shows the significant very low birth weight clusters that were defined by the *SaTScan* software. The most highly significant cluster occurs in the metropolitan Denver area, with another cluster in the Colorado Springs-Pueblo area.

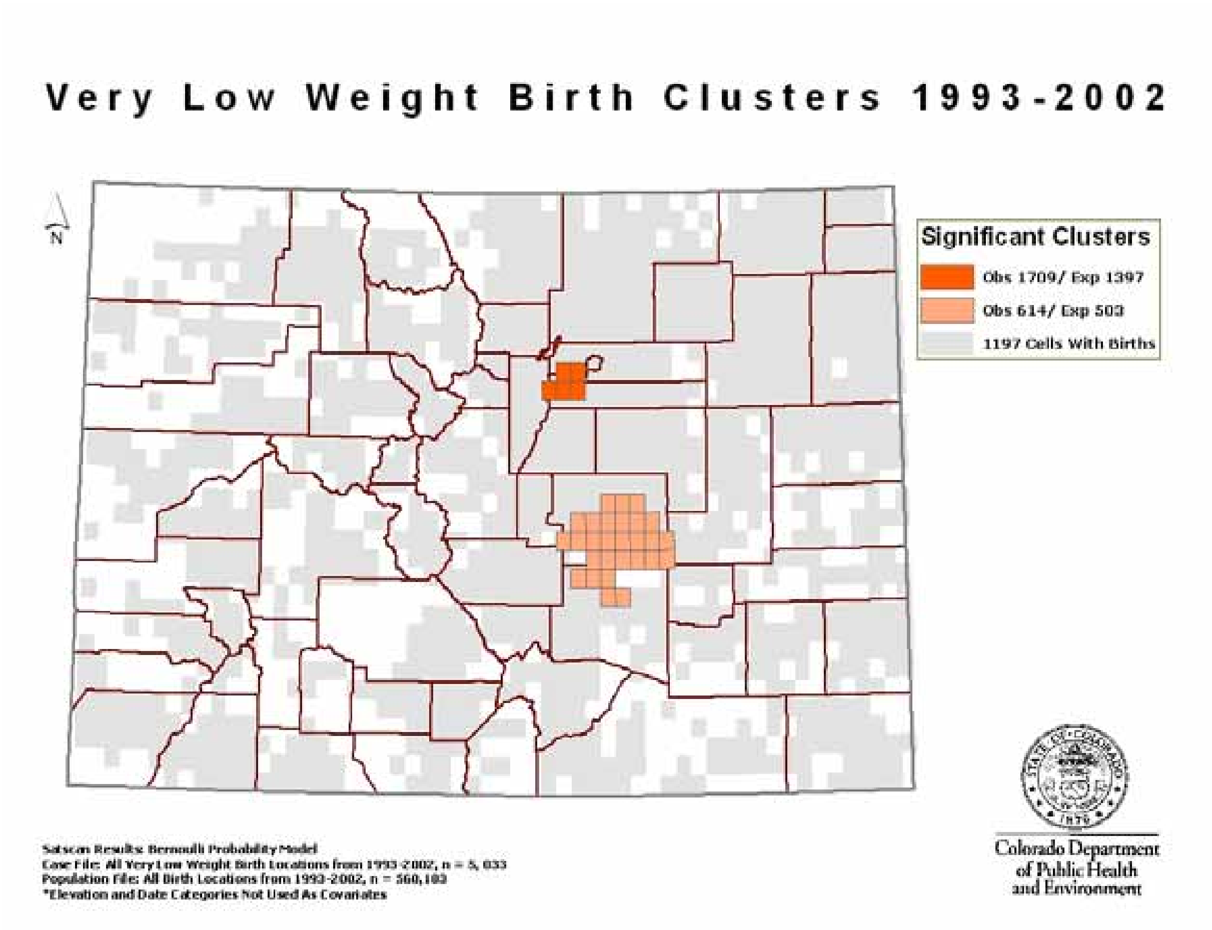

Figure 5. Very Low Weight Birth (<1500 grams) Clusters in Colorado (1993-2002)

The next map examines smoking patterns of mothers (normal and low birth weights) across the state without regard to elevation, so these clusters may or may not contain similar elevations. Several significant clusters were identified, the largest covering a large area of south central and southwestern Colorado. This area contains varying altitude, while the other clusters seem to be in areas of more uniform altitude.

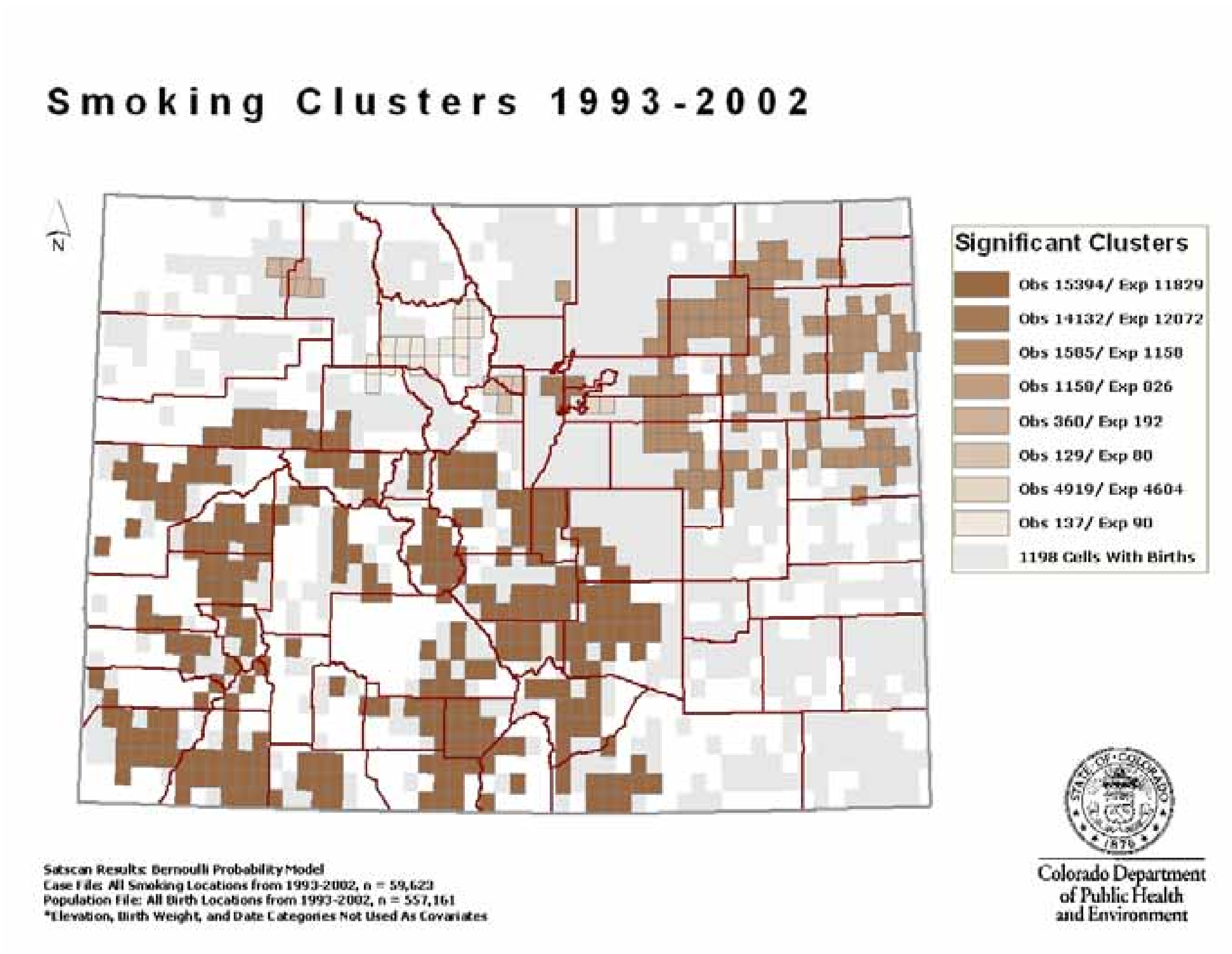

Figure 6. Smoking Clusters in Colorado

Our last map displays clusters of smoking and low birth weights, and may be more useful to target specific areas for interaction. Five statistically significant clusters of low birth weight associated with maternal smoking are identified. One cluster is located in metropolitan Denver, three overlay mountain ranges and high plateaus in the central and western portion of the state, and one is located on the eastern plains. These clusters somewhat model the smoking clusters.

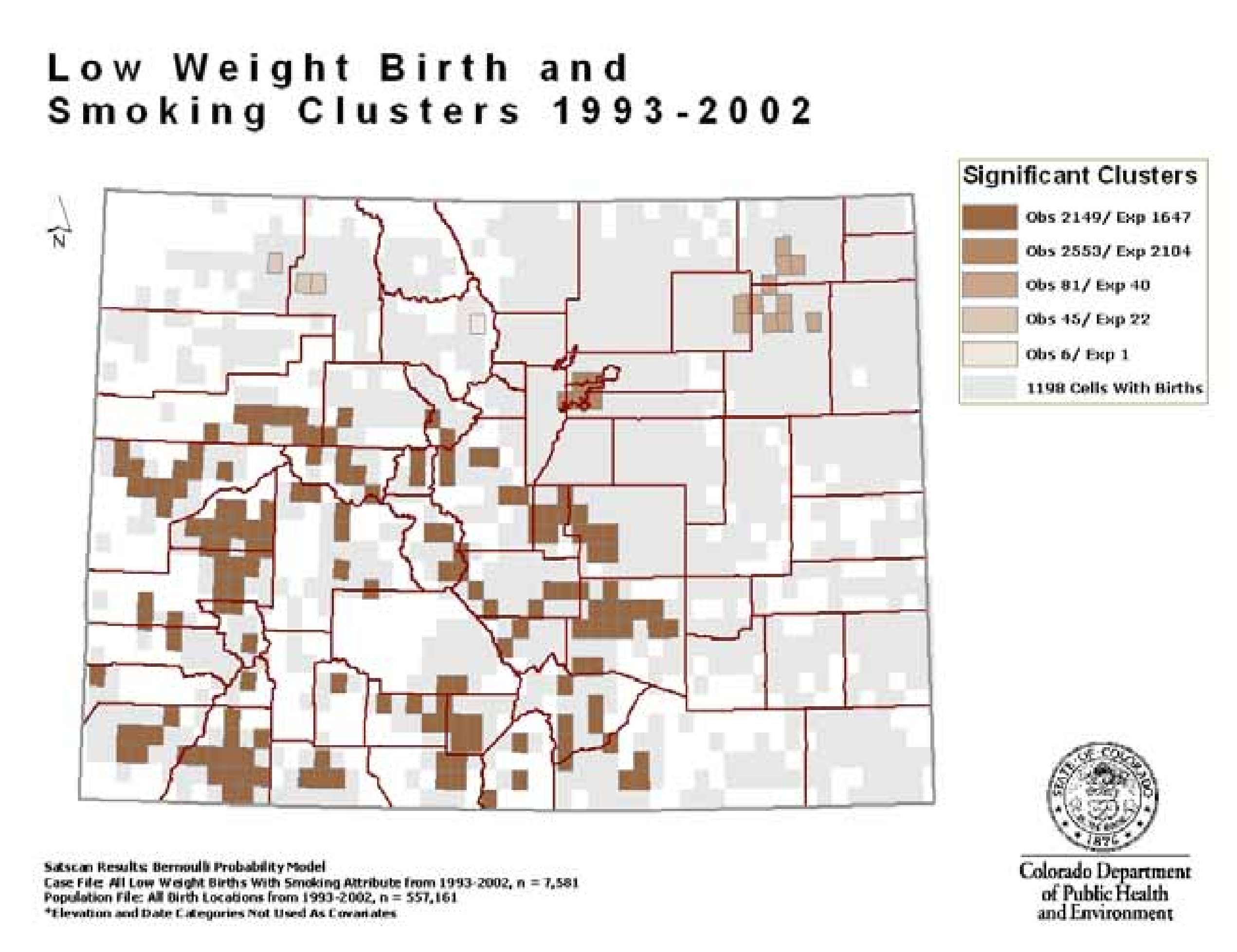

Figure 7. Smoking Clusters in Colorado

Summary

In the past, epidemiologists have been limited to describing or exploring public health issues in non-spatial terms. The use of GIS with statistical analysis software can paint a detailed picture of issues such as low birth weight and altitude, pinpoint locations of incidence, and allow for powerful analysis of other demographic characteristics that may contribute to risk. The spatial analysis of low birth weights in Colorado also demonstrated an effective methodology for documenting, geo-referencing, and utilizing a large number of public health records, in an environment that can be repeated on other public health data sets. The next step in our low birth weight analysis is to maintain this data in a secure yet accessible environment, continue to build communication lines with local epidemiologists and public health officials in areas where there are significant clusters of smoking and low birth weight, and build a coalition of interest in this type of visual analysis of public health data. One way of presenting information about public health issues and historical trends is to develop a “Public Health Atlas,” utilizing interactive and online mapping. CDPHE is currently exploring ways to develop an atlas for use by the general public and public health officials.