CONF-950168--2-REV.2

LBL-36630 Rev. 2
UC-405

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

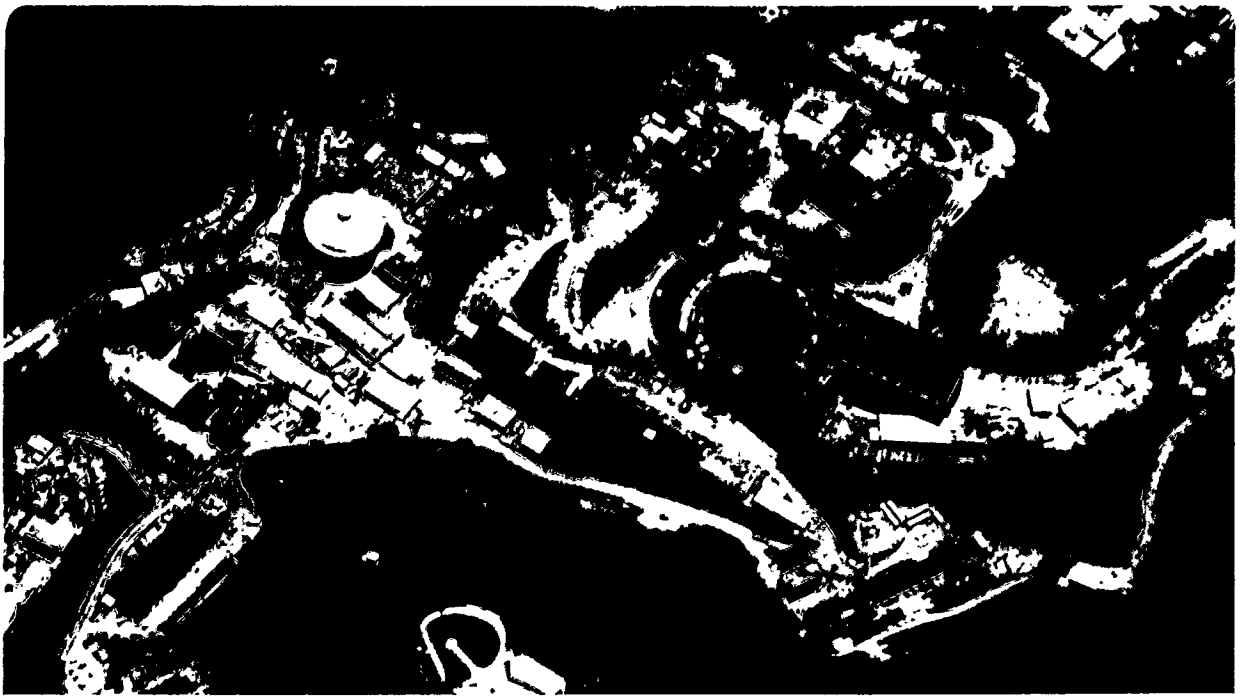
Information and Computing Sciences Division

Presented at the Symposium on Statistical Methods, Atlanta, Georgia, January 24–26, 1995, and to be published in the Proceedings

Use of Density Equalizing Map Projections (DEMP) in the Analysis of Childhood Cancer in Four California Counties

D.W. Merrill, S. Selvin, E.R. Close, and H.H. Holmes

October 1995



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

LBL-36630 Rev.2
UC-405

USE OF DENSITY EQUALIZING MAP PROJECTIONS (DEMP) IN THE ANALYSIS OF CHILDHOOD CANCER IN FOUR CALIFORNIA COUNTIES

D.W. Merrill[a], S, Selvin[b], E.R. Close[c] and H.H. Holmes[d]

Information and Computing Sciences Division

Ernest Orlando Lawrence Berkeley National Laboratory

University of California

Berkeley, CA 94720

October 1995

a. Information and Computing Sciences Division (ICSD), Lawrence Berkeley National Laboratory (LBNL), One Cyclotron Road, Berkeley CA 94720. Mail stop 50B-2239. Tel: 510-486-5063. dwmerrill@lbl.gov.
b. School of Public Health, 140 Warren Hall, University of California, Berkeley CA 94720. Tel: 510-642-4618. selvin@psi.berkeley.edu.
c. 568 Arlington Avenue, Berkeley CA 94707. Tel: 510-525-2984. elonclose@aol.com.
d. ICSD, LBNL Mail stop 50F. Tel: 510-486-5742. hhholmes@csa.lbl.gov.

This work was supported by the Director, Office of Epidemiologic Studies; Office of Health Studies; Office of Environment, Safety and Health; U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

USE OF DENSITY EQUALIZING MAP PROJECTIONS (DEMP) IN THE ANALYSIS OF CHILDHOOD CANCER IN FOUR CALIFORNIA COUNTIES

SUMMARY

In studying geographic disease distributions, one normally compares rates among arbitrarily defined geographic subareas (e.g. census tracts), thereby sacrificing the geographic detail of the original data. The sparser the data, the larger the subareas must be in order to calculate stable rates. This dilemma is avoided with the technique of Density Equalizing Map Projections (DEMP). Boundaries of geographic subregions are adjusted to equalize population density over the entire study area. Case locations plotted on the transformed map should have a uniform distribution if the underlying disease rates are constant.

The present report describes the application of the DEMP technique to 401 childhood cancer cases occurring between 1980 and 1988 in four California counties, with the use of map files and population data for the 262 tracts of the 1980 Census. A k'th nearest neighbor analysis provides strong evidence for geographic non-uniformity in tract rates ($p < 10^{-4}$). No such effect is observed for artificial cases generated under the assumption of constant rates.

Work is in progress to repeat the analysis with improved population estimates derived from both 1980 and 1990 Census data. Final epidemiologic conclusions will be reported when that analysis is complete.

INTRODUCTION

Density equalizing map projections (DEMP), also known as population maps, cartograms or anamorphoses, have long been used for display of thematic data.[1-3] The DEMP technique is appropriate for analyzing disease distributions because on a density equalized map, population density is constant. Therefore cases should occur randomly and uniformly under the null hypothesis of equal risk.

Normally, one analyzes geographic disease distributions by comparing rates among different subareas. Relative to conventional methods, the DEMP technique has several advantages:

- A density equalized map portrays both individual events and rates, and can be understood by untrained observers. The full geographic detail of the data is preserved, and arbitrary grouping of subareas is avoided.
- The DEMP technique avoids the calculation of unstable rates. The technique is appropriate, and even works best, when the number of cases is small.
- No *a priori* knowledge is required for testing the null hypothesis of equal risk. Hence the DEMP technique is appropriate for automatic analysis of routinely collected surveillance data. The uniformity of the density equalized map can be tested with simple statistical techniques.

DEMP ALGORITHM

Development of a practical DEMP algorithm has been a goal of the Lawrence Berkeley Laboratory (LBNL) Populations at Risk to Environmental Pollution (PAREP) project since 1985. Computer algorithms developed elsewhere are suitable for graphic display purposes, but not for the requirements of statistical analysis. Earlier LBNL algorithms addressed this problem but were too slow for practical use.[4,5] In 1993 Gusein-Zade and Tikunov described a new algorithm[6] which was independently implemented and tested in 1994 at LBNL.[7] New features were added in January 1995.[8]

Required input data for the DEMP program are (1) locations of the individual cases (2) latitude/longitude coordinates of the boundary points of the tracts, which are represented as polygons (3) estimates of the population at risk in each census tract. The computer program adjusts the boundaries of all the census tracts so that the transformed areas are exactly proportional to population at risk. In the transformation, the original adjacencies among the tracts are preserved, and distortion of tract shapes is kept to a minimum.

The case locations can be specified either by latitude and longitude, or by the tract to which they belong. If exact case locations are used, they are included in the map prior to density equalization and transformed in the same way as the boundary points. If only the census tract of residence is known, each case can be plotted at a random point within its tract, either before or after density equalization.

The 'push' to be applied to each point in the map, which can be calculated in closed form, is a complex sum of line integrals over all the line segments in all the polygon

boundaries. The transformation is performed in a number of small iterations, to ensure that polygon boundaries do not self-intersect during the transformation. At each step the present area of each polygon is compared with its target area, which is proportional to the population at risk. The 'pushes' generated by a given polygon shrink to zero as its area approaches the target value; the iteration process is terminated when the all the polygon areas have converged to their target values. Depending on the non-uniformity of population density in the original map, about ten iterations are generally sufficient to make the residual error negligibly small.

FOUR-COUNTY CHILDHOOD CANCER STUDY

The present study analyzes 401 childhood cancer cases occurring between 1980 and 1988, in children of ages 0 through 14, in four California counties: Fresno, Kings, Kern and Tulare. The data were originally collected and analyzed by the California State Department of Health Services (DHS) to investigate a reported childhood cancer cluster in McFarland, California.[9,10]

DHS subdivided the four counties into 101 communities. Six communities had rates that fell outside 95% confidence limits (three with more cases than expected and three with fewer cases than expected). The result is consistent with the null hypothesis of uniform underlying rates. One community (McFarland) had an elevated rate outside the 99% confidence limit, exactly what would have been expected from chance alone.

The same data were re-analyzed at LBNL as follows. Rather than the 101 communities of the DHS analysis, LBNL used the 262 tracts defined in the 1980 Census. Census tracts

are preferable to communities because they provide greater geographic detail, are more nearly equal in population, and are more homogeneous with respect to socio-economic characteristics.

DATA PREPARATION

Case Locations

DHS kindly provided to LBNL the case records from the earlier analysis.[10] Each case record includes the latitude/longitude coordinates of the location of residence. As in the DHS analysis, LBNL analyzed all 401 cases as a single data set, without regard to year of diagnosis, age category, gender or race.

Census Tract Boundaries

The 1980 Census tract map file in LBNL SEEDIS (Socio-Economic Environmental Demographic Information System)[11] was used, which was originally purchased from National Planning Data Corporation. To reduce computing time in the DEMP process, the map was reduced by successively removing points from the tract boundaries until the smallest remaining details were no smaller than 20% of the area of their respective tracts. 'Doughnut' tracts were converted to simple polygons by 'cutting' one side of the doughnut; i.e., introducing an artificial segment connecting the external boundary polygon with the internal 'hole' polygon. The resulting map has 264 simple polygons including 262 tracts and two very small lakes whose area shrank to zero in the density equalization.

To prevent the polygon boundaries from self-intersecting during the DEMP process, the 264 polygons were subdivided into 1211 triangles, which are the subareas which were

density equalized. We chose the unique Delaunay triangulation, which divides each polygon into triangles that are as nearly equiangular as possible.[12] Each polygon was triangulated by inspecting its boundary points one at a time. If a candidate point and its two adjacent points satisfied the Delaunay criterion (below), the triangle formed by those three points was removed and the process was reapplied to the remaining polygon, which now has one less boundary point than before. The process was repeated for each polygon until only triangles remained.

A triangle satisfies the Delaunay criterion if and only if the circle drawn through its three vertices contains no other points of the polygon boundary. Regardless of the order in which Delaunay triangles are removed, the triangulation of a given polygon is unique (except in the rare situation where four adjacent boundary points exactly form a rectangle, in which case two choices are possible).

It was discovered that a large map composed entirely of triangles cannot be density equalized. In principle there are enough degrees of freedom, but in practice a solution was never reached. The problem was solved by dividing each triangle boundary segment into two segments, and allowing the positions of the break points to vary freely during the DEMP process.

Census Tract Populations

The DHS analysis was performed before the 1990 Census data became available. It used 1980 Census data, plus intercensal estimates from the California Department of Finance.

Populations were estimated for 101 communities rather than 262 tracts. Unfortunately the DHS population estimates were inadvertently erased after the analysis was complete.

In the analysis described here, LBNL used the 1980 Census tract populations for children of ages 0-17, which are readily available in SEEDIS. The most important task, which had never succeeded previously, was to test the feasibility of the DEMP algorithm in a large study area with highly non-uniform population density.

An improved future analysis will use instead the 1980-88 population at risk for children of ages 0-14. Such a calculation is tedious but straightforward; it requires both 1980 and 1990 Census data map files, and must account for intercensal changes in tract boundaries. The revised analysis is in progress and will be reported at a later date, along with epidemiologic conclusions.

DENSITY EQUALIZATION

The 262 tract boundaries (solid), the 1211 triangle boundaries (dotted), and the 401 case locations are shown in Figure 1, prior to density equalization. Target populations were assigned to each of the 1211 triangles by assuming uniform population density within each tract. The map with 1211 triangles was density equalized in ten equal steps (*i.e.*, with step size in each iteration equal to 1/10, 1/9, ... 1/2, 1/1 of the 'remaining distance'). The density equalization required about 20 hours on a Sun SPARC 10 work station. Because a break point had been introduced into each triangle boundary segment, the 1211 triangles gradually assumed the form of hexagons as the density equalization progressed.

Figure 1. Filtered and triangulated map, with 401 cases.

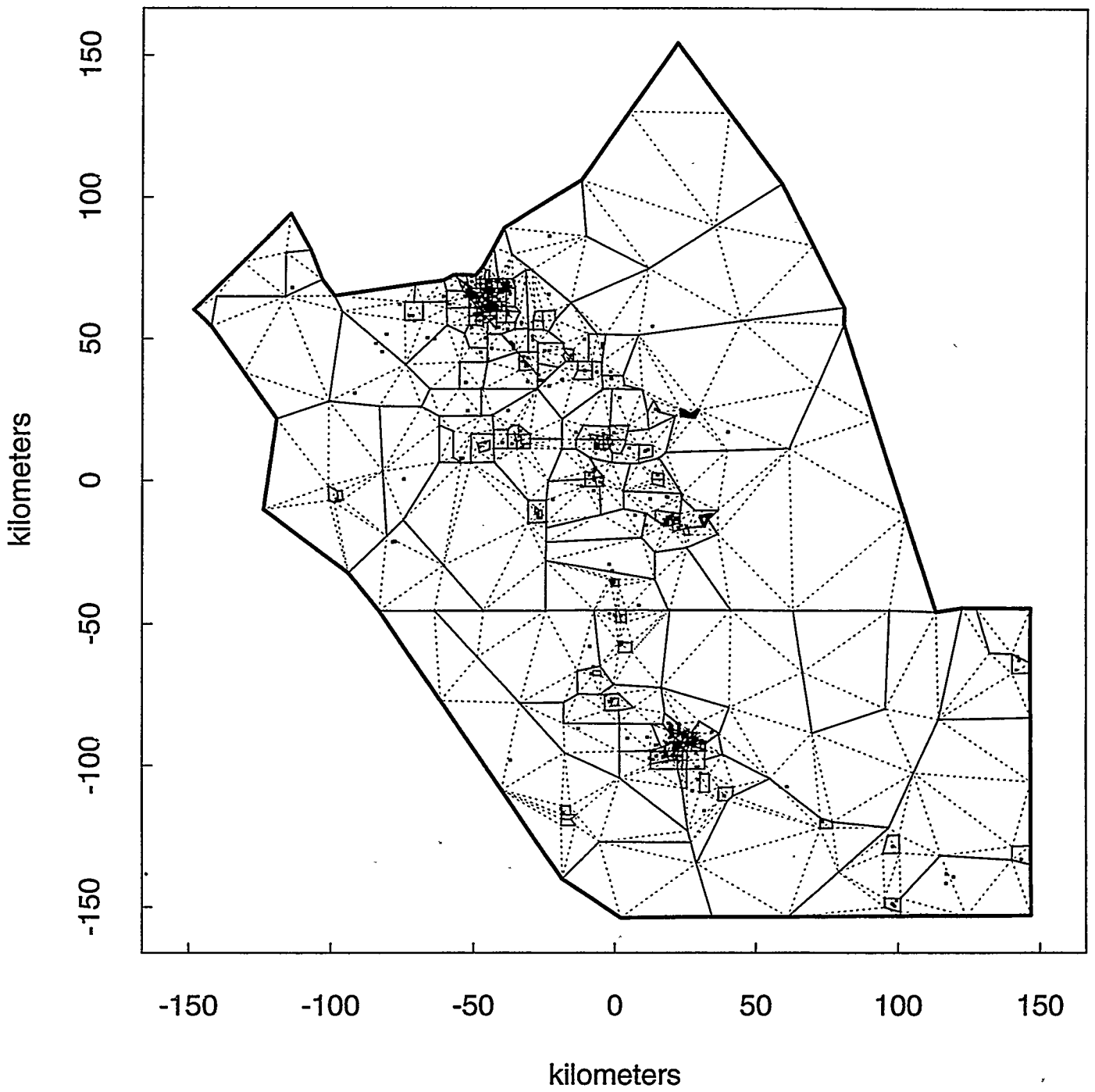


Figure 2. Tracts, hexagons and 401 cases, density equalized.

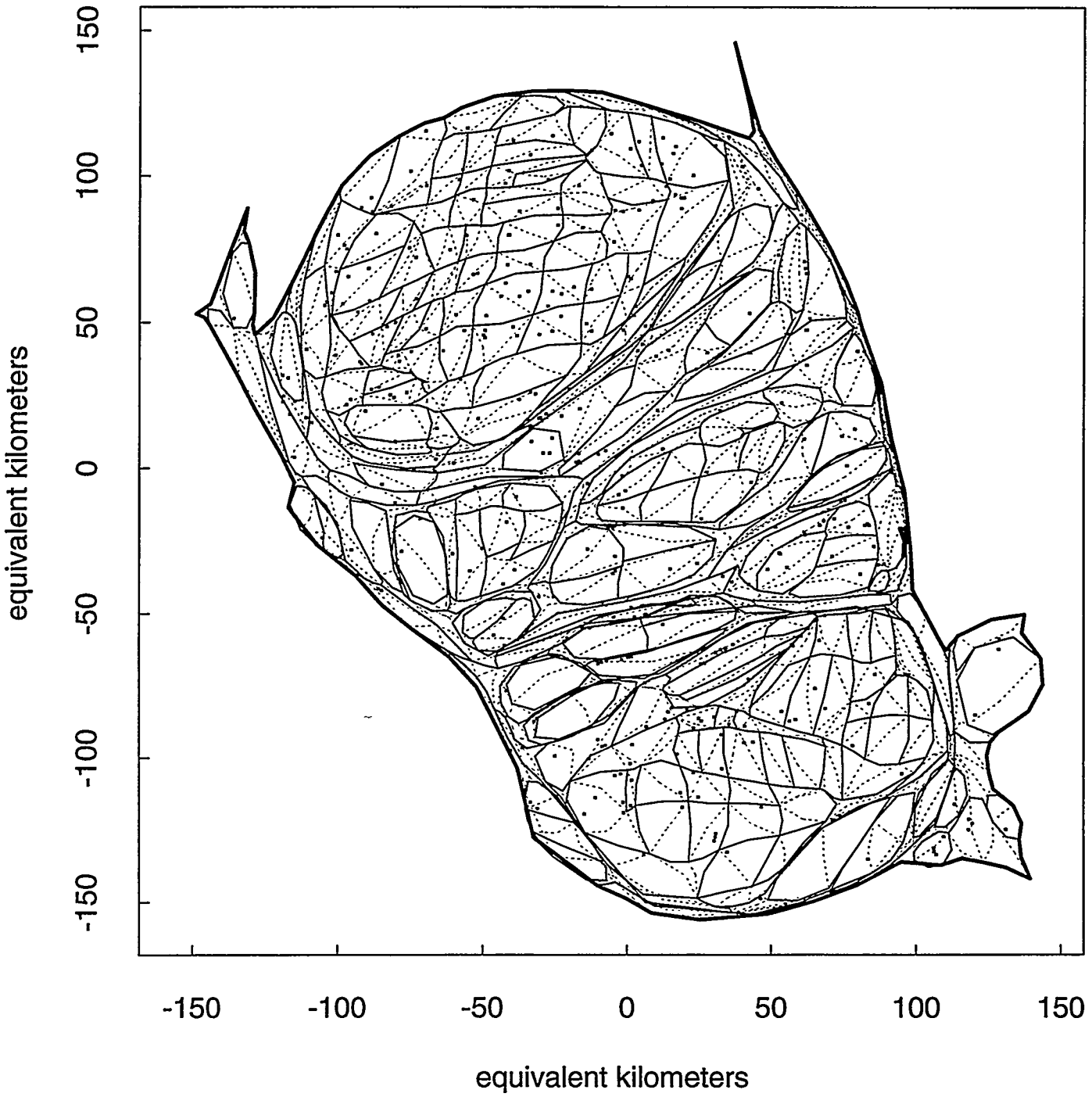


Figure 2 shows the tract boundaries (solid), the hexagon boundaries (dotted), and the locations of the 401 cases after density equalization.

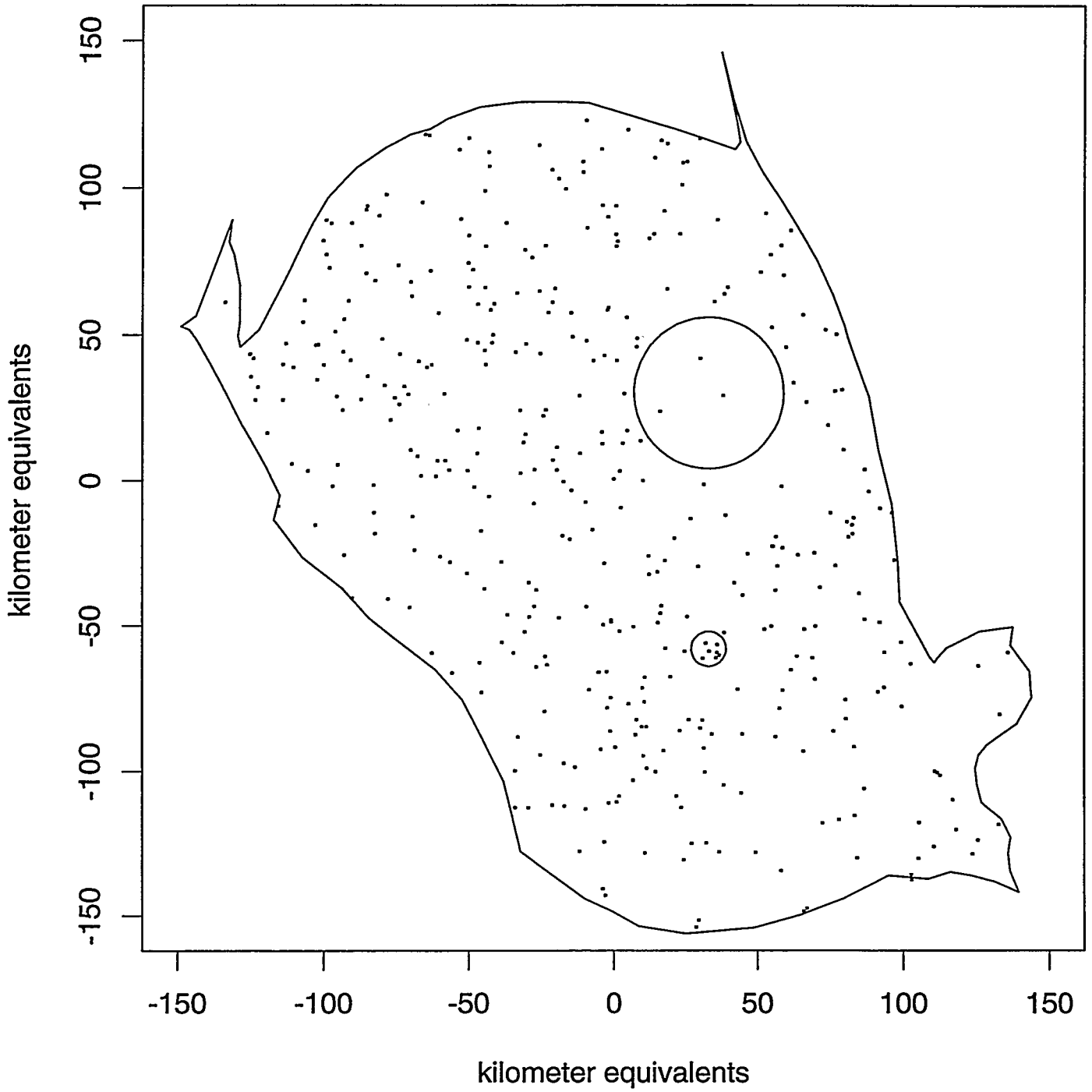
The same transformation was also applied to artificial cases which were randomly generated under the assumption of equal risk. Statistical tests and visual analysis (not shown here) showed that the artificial cases are uniformly distributed on the density equalized map. These tests confirmed that the DEMP transformation does not introduce artificial clustering.[8]

The same transformation was also applied to the 401 real cases plotted not at their actual locations, but at random locations within their respective tracts. This is the analysis that is appropriate if only the census tract of residence (and not the actual location) is known. This procedure is also required for the visual analysis and statistical analysis described below. Geographic detail below the tract level must be removed from the case data, to correspond to the tract level population and map data; otherwise clusters of cases remain that may be due to non-uniform population density within individual tracts. To equalize population density *within* tracts requires population data and map files with greater geographic detail.

VISUAL ANALYSIS

The density equalized map, with internal boundaries removed, is shown in Figure 3. Here, unlike Figure 2, each case was randomly plotted within its own tract prior to density equalization. Figure 3 portrays visually not only the locations of individual cases but also the geographic variation of rates. Even an untrained observer can quickly distinguish relatively dense or sparse areas, which may suggest hypotheses for further investigation.

Figure 3. Density equalized map, with 401 cases.



STATISTICAL ANALYSIS

The small circle in Figure 3, a relatively dense region, contains seven cases in an area where only 0.9 cases are expected. The large circle, a relatively sparse region, contains only three cases in an area where 17 cases are expected. The corresponding Poisson-based p-values (4.3×10^{-5} and 4.5×10^{-5} for the small and large circles respectively) represent the probability of finding at least this many (or at least this few) cases in a *specific* region. Such a calculation is not appropriate in the present analysis, since the DHS analysis was investigated *in response to* a reported cluster of cases in McFarland. Random variation will always produce clusters *somewhere* in any study area.

Various authors, for example Cuzick and Edwards[13] and Kulldorff and Nagarwalla[14], have proposed methods for evaluating the statistical significance of a cluster detected *anywhere* in the study area. Although such techniques do not required a density-equalized map, the underlying theory is greatly simplified if the expected distribution under the null hypothesis is uniform.

K'TH NEAREST NEIGHBOR ANALYSIS

We now present an example of a simple analysis where a density-equalized map is *required*; namely, k'th nearest neighbor. The analysis presented here is a generalization of a nearest neighbor analysis ($k=1$) presented in an earlier report.[8] The k'th nearest neighbor analysis simply measures overall spatial nonuniformity; it does not identify the location or the significance of individual clusters.

On the density equalized map, each of the n (= 401) cases has a nearest neighbor distance which is the distance to the nearest of all the n-1 other cases. To test for spatial randomness among the n cases, we calculate the mean nearest neighbor distance d_1, which is the average of these n distances. In the k'th nearest neighbor analysis we consider not only the distance d_1 to the nearest neighbor (k=1), but also the distance d_2 to the next nearest (k=2), and d_k for all values of k up to n-1. The different values of k are not independent tests, but different choices of a variable parameter. The analysis is sensitive to either small-area or large-area effects, depending on the value of k.

The expected value of d_k (and the standard error and all higher moments) can be derived under the assumption that points are spatially distributed at random. The formula for the expected value of d_k is given by Cressie.[15] For comparison with theory it is convenient to make d_k dimensionless; that is, to express d_k in units which are equal to the square root of A/n, where A is the area of the region containing n cases. With this convention the expected value of d_1 (for example) is 0.5 for points distributed at random. An observed value of d_1 less than 0.5 indicates mutual attraction, or clustering; a value greater than 0.5 indicates mutual repulsion, or anti-clustering.

Boundary Bias

For any finite study area, the observed value of d_k is biased upward relative to the theoretical value. This occurs because cases near the boundary of the study area have reduced probability of having close neighbors, and so their k'th nearest neighbor distances are biased upward. The boundary bias becomes increasingly important as k increases.

As n increases the bias in d_k becomes smaller, but so does the standard error of d_k. At least for $k=1$, it happens that the bias and the standard error of d_1 have the same functional dependence on n, so that the ratio of the bias to the standard error is independent of n. For the area under investigation, that ratio is approximately 1; that is, for $k=1$ and any value of n the boundary bias of d_1 is equivalent to a one standard deviation effect.[8]

In comparing d_k with the theoretical value, it is necessary to correct for the boundary bias. One simple method is to generate artificial random cases outside the boundary of the density equalized map, with the same density as that observed inside. The external cases are included as nearest neighbor candidates when calculating d_k for the cases inside the boundary. With the external random cases included, k is not limited to n-1 and can be as large as desired provided the extended area is sufficiently large. For very large k (about 2000 in our analysis), the k'th nearest neighbor of *every* case is a random external point, so no clustering effects are observed for k greater than this value.

Analysis of Case Locations on the Density Equalized Map

In the present analysis, the same DEMP transformation was simultaneously applied to three separate sets of points: (a) 401 artificial case locations randomly generated assuming equal risk (b) the 401 actual case locations (c) the 401 cases plotted at random locations within their respective tracts. For correction of the boundary bias, each data set was augmented (after density equalization) with a set of random points outside the boundary. To reduce the effects of random variability, all three data sets (a), (b) and (c) were replicated 20 times. The random points in all 60 samples were completely independent of each other.

In Figure 4, as a function of k, we present the ratio of the observed mean d_k to the theoretical mean for samples (a), (b), and (c).

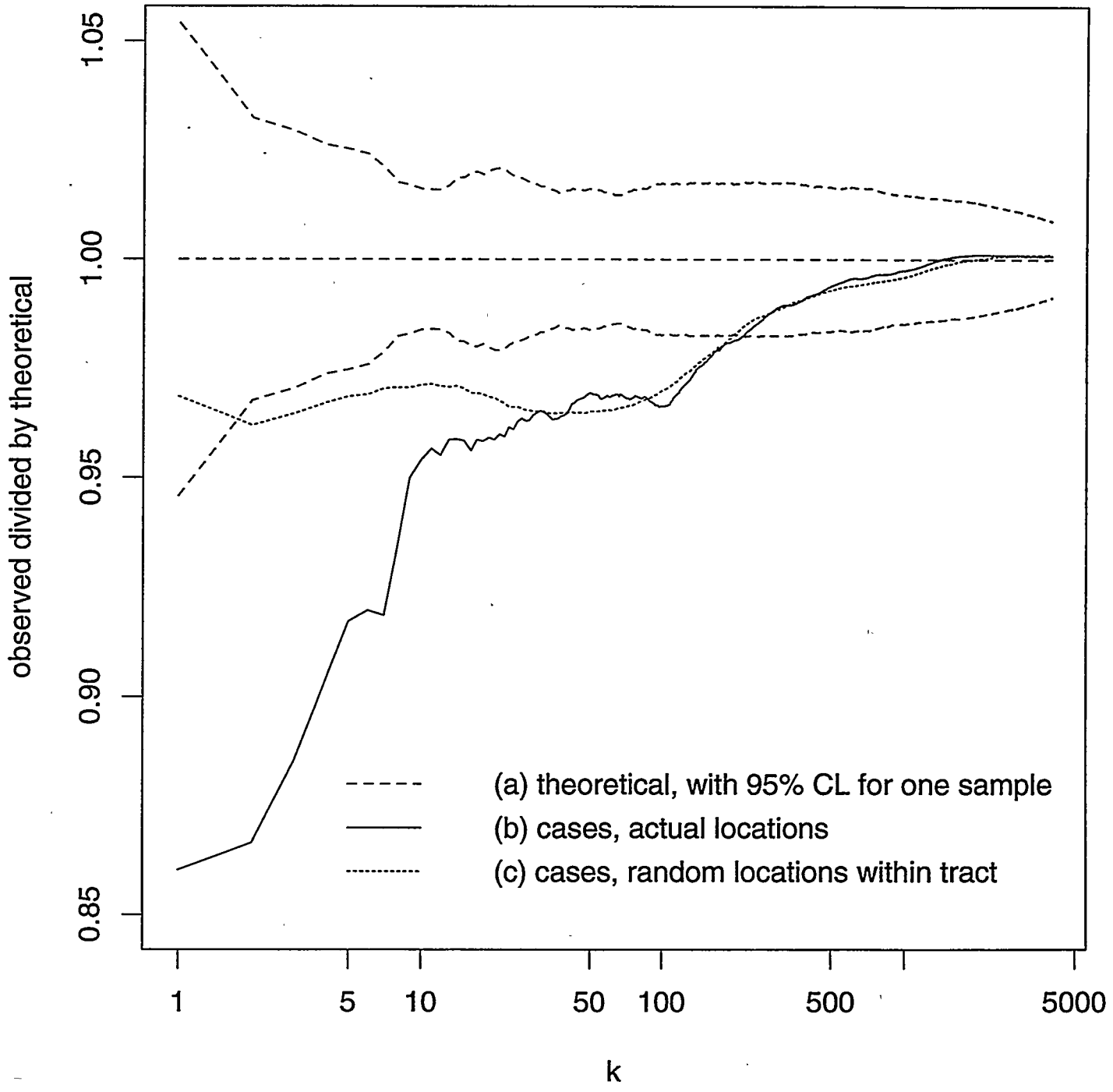
Sample (a): Random Artificial Cases

The straight dashed line at y=1.00 represents the expected value of random artificial cases generated under the null hypothesis. The symmetric dashed curves above and below the straight line represent the 95% confidence interval (1.96 s.d.) of variability for a single sample, which was estimated empirically from the variance of 20 independent samples.[16] At k=1, the single-sample (1.96 s.d.) confidence interval is (0.94, 1.06). For all values of k, the 20-sample mean value of d_k was found to be consistent with the theoretical value. This test indicates that the case simulation and boundary correction procedures are performing as expected.

Sample (b): Actual Case Locations

The solid curve in Figure 4 represents the observed value of d_k relative to the theoretical value, for the actual case locations. Twenty samples (b) were analyzed, which differ only in the random external cases required for correction of the boundary bias. The value at k=1 is 0.85, corresponding to an observed clustering effect of -4.9 s.d. ($p < 10^{-6}$). The significant clustering at small values of k is due at least in part to non-uniform population density within individual tracts, which cannot be equalized with the available map files and population data.

Figure 4. Mean k'th nearest neighbor distance.



Sample (c): Actual Cases, Plotted at Random Locations within Tract

The lightly dotted curve in Figure 4 represents the observed value of d_k relative to the theoretical value, for the actual cases plotted at 20 different random locations within their respective tracts. The value at k=1 is 0.96, corresponding to a clustering effect of only -1.3 s.d. (p = 0.10). Not surprisingly, clustering is insignificant for k=1 because the nearest (k=1) neighbor of a case is frequently in the same tract as the case, and in sample (c) the case locations within each tract have been randomized.

As k increases, the distance from a case to its k'th nearest neighbor becomes large compared with the dimensions of a single tract, so the randomization in (c) becomes insignificant. For values of k greater than about 40, the values of d_k from samples (b) and (c) differ only due to random variation. For values of k greater than about 2000, only the randomly generated cases outside the boundary contribute to d_k, so samples (a), (b), and (c) are all in agreement within statistical variation.

The most interesting finding is that sample (c), for any value of k between 2 and about 200, shows evidence of significant clustering ($p < .025$). This effect, unlike that in (b), is due entirely to observed rate differences among different tracts. The effect is most marked for k between 20 and 100, where the evidence for non-uniformity of tract rates is about four standard deviations ($p < 10^{-4}$).

DISCUSSION

The LBNL analysis of the four county data set provides strong statistical evidence ($p < 10^{-4}$) for non-uniformity of tract rates of childhood cancer. However, the observed

effects may be entirely due to biases in the input data. The use of 1980 population data in conjunction with 1980-1988 case data is an important defect that must be remedied before epidemiologic conclusions can be discussed. It is very likely that different tracts experienced different rates of population growth after 1980, which could dramatically affect the results presented here.

In addition, confounding bias can result from differential census undercount of certain social groups (*e.g.* Hispanics or migrant workers) coupled with non-uniform geographic distribution of those groups. To some extent this can be checked by stratifying the analysis on those social characteristics.

Finally, errors in the input data can produce statistically significant effects if those errors have any cause other than random statistical variation. Evidence of geographic nonuniformity of rates, however statistically significant, is not meaningful unless a systematic pattern is detected and unless all sources of bias can be ruled out.

The caveats just stated concern the validity not of the DEMP methodology but rather of the input data. These caveats would apply equally to any analysis of the same data.

CONCLUSIONS

A major accomplishment described in this report is the successful density equalization (in Figure 2) of the complex and highly non-uniform map in Figure 1. Analysis of artificially generated cases showed that significant biases are not introduced by the density equalization.[8] Figure 3 portrays individual cases and rates in a single display that can be understood by untrained observers.

Various simple statistical techniques can be applied to density equalized maps. One example, the mean k'th nearest neighbor distance, provides a sensitive and unbiased measure of overall non-uniformity of rates, provided one corrects for boundary effects and for non-uniform population distribution within individual tracts. Figure 4 presents evidence of geographic nonuniformity of tract rates, with statistical significance equivalent to four standard deviations ($p < 10^{-4}$).

Application of the DEMP technique requires significant computing resources. Opportunities exist for an improved implementation, now that the feasibility of the algorithm has been demonstrated. The DEMP technique can become a valuable tool for routine surveillance activities, especially if automatically coupled to data bases containing the necessary population data and map files for all regions of the United States.

The analysis presented here used readily available 1980 Census data. Calculation of improved population estimates will require map files and population data from both the 1980 and 1990 Censuses. The calculation is straightforward but tedious due to intercensal changes in census tract boundaries. Reanalysis of the four-county data set with improved population estimates is in progress; final epidemiologic conclusions will be presented when that analysis is complete.

ACKNOWLEDGMENTS

The authors thank the California Department of Health Services for providing the case data from the Four County Childhood Cancer Study. Raymond Neutra, Peggy Reynolds, and Enid Satariano provided useful guidance and assistance.

We have enjoyed an active electronic correspondence with Vladimir Tikunov and Sabir Gusein-Zade, the Russian authors of the DEMP algorithm we have implemented and modified. Michael Mohr wrote many of the programs used to manipulate map files. Early programmatic support was provided through the efforts of Carl Quong and Robert Goldsmith.

This work was supported by the Director, Office of Epidemiologic Studies; Office of Health; Office of Environment, Safety and Health; U.S. Department of Energy under Contract No. DE-AC03-76-SF00098.

REFERENCES

The present report is a condensed and revised version of earlier unpublished reports (Refs. 7, 8, 16, and 17). Electronic versions of those reports are available on the World Wide Web (WWW) at the Uniform Resource Locators (URL's) indicated below.

1. Wallace, J. W. 'Population map for health officers'. *American Journal of Public Health*, **16,** 1023 (1926).
2. Gillihan, A. F. 'Population Maps'. *American Journal of Public Health*, **17,** 316-319 (1927).
3. Raisz, E. 'The rectangular statistical cartogram'. *Geographical Review*, **24,** 292-296 (1934).

4 Schulman, J., Selvin, S. and Merrill, D. W. 'Density equalized map projections: a method for analyzing clustering around a fixed point'. *Statistics in Medicine*, **7**, 491-505 (1988).

5. Merrill, D. W., Selvin, S. and Mohr, M. S. 'Analyzing Geographic Clustered Response'. In 1991 Proceedings of the Section on Statistics and the Environment of the American Statistical Association, Atlanta GA, August 18-22, 1991.

6. Gusein-Zade, S. M. and Tikunov, V. S. 'A New Technique for Constructing Continuous Cartograms'. *Cartography and Geographic Information Systems*, **20(3)**, 167-173 (1993).

7. Close, E. R., Merrill, D. W. and Holmes, H. H. 'Implementation of a new algorithm for Density Equalizing Map Projections (DEMP)'. Report LBL-35738, July 1995 (134 pages). WWW URL: http://cedr.lbl.gov/pdocs/tr940401/all.html.

8. Merrill, D., Selvin, S., Close, E. R. and Holmes, H. H. 'Use of density equalizing map projections (DEMP) in the analysis of childhood cancer in four California counties'. Report LBL-36630, January 1995 (74 pages). WWW URL: http://cedr.lbl.gov/pdocs/cdc9501/lbl36630.html.

9. Satariano, E., Reynolds, P., Smith, D., and Goldman, L. 'The Four County Study of childhood cancer incidence'. Interim report I. Environmental Epidemiology and Toxicology Branch, California Department of Health Services, May 1990.

10. Reynolds, P., Satariano, E. and Smith, D. 'The Four County Study of Childhood cancer incidence'. Interim report II. Environmental Epidemiology and Toxicology Program, California Department of Health Services, October 1991.

11. Merrill, D. 'The Lawrence Berkeley Laboratory (LBNL) Socio-Economic Environmental Demographic Information System (SEEDIS)'. WWW URL: http://cedr.lbl.gov/mdocs/seedis/seedis.html (1995).

12. Boots, B. N. 'Voronoi (Thiessen) Polygons'. In series: *Concepts and Techniques in Modern Geography (CATMOG)*, Geo Books, Norwich UK (1987).

13. Cuzick, J. and Edwards, R. 'Spatial clustering for inhomogeneous populations'. *Journal of the Royal Statistical Society, Series B* **52,** 73-104 (1990).

14. Kulldorff, M. and Nagarwalla, N. 'Spatial Disease Clusters: Detection and Inference'. *Statistics in Medicine,* **14,** 799-810 (1995).

15. Cressie, N. *Statistics for Spatial Data.* John Wiley & Sons, Inc., New York, 1991.

16. Merrill, D. W., Selvin, S. and Close E.R. 'Use of density equalizing map projections (DEMP) in the analysis of childhood cancer in four California counties'. Abstract in proceedings of the 1995 CDC/ATSDR Symposium on Statistical Methods; Atlanta GA, January 25-26, 1995. Graphics (25 pages) are at WWW URL: http://cedr.lbl.gov/pdocs/cdc9501/25graphics.html.

17. Merrill, D., Selvin, S., Close, E. R. and Holmes, H. H. 'Use of density equalizing map projections (DEMP) in the analysis of childhood cancer in four California counties'. Report LBL-36630 Rev., April 1995 (18 pages). WWW URL: http://cedr.lbl.gov/pdocs/cdc9504/cdc9504.html.

FIGURE CAPTIONS

1. Filtered and triangulated map, with 401 cases. Unnecessary geographic detail has been removed, and the 262 tracts (solid lines) have been subdivided into 1211 triangles (dotted lines). Locations of the 401 cases are indicated by dots.
2. Tracts, hexagons and 401 cases, density equalized. This is the same map as in Figure 1, after density equalization. Prior to density equalization, the 1211 triangles were converted to hexagons by dividing each triangle boundary segment into two segments.
3. Density equalized map, with 401 cases. Prior to density equalization, each case was plotted at a random location within its own tract. The small circle contains seven cases in an area where only 0.9 cases are expected. The large circle contains only three cases in an area where 17 cases are expected.
4. Mean k'th nearest neighbor distance d_k, divided by the theoretical value, plotted as a function of k. The symmetric dashed curves (a) represent the 95% (1.96 s.d.) confidence interval for a single sample, about the theoretical value of 1.00. The confidence interval was calculated empirically from 20 samples of artificial cases, randomly generated under the null hypothesis of equal risk. The solid curve (b) was obtained from the actual case locations; the lightly dotted curve (c) was obtained from the actual cases plotted at random locations within their respective tracts. For values of k between 20 and 100, the evidence for non-uniformity of tract rates in (c) is about four standard deviations ($p < 10^{-4}$).

```
document location (Word Perfect 5.1)
C:\docs\parep\cdc9510\  =
    csr6.lbl.gov:c:\docs\parep\cdc9510\
      cdc9510.22          10/10/95
      simdm.wp            10/10/95
K:\docs\parep\cdc9510\  =
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      cdc9510.22          10/10/95
      simdm.wp            10/10/05

document location (HTML)
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      cdc9510.html        10/10/95


figure locations (Splus)
Figure 1:
    parep2.lbl.gov:/export/home/u0/cedrdv/merrill/public_html/figs/
      fig25a.csh          3/24/95
      fig25a.ps           3/24/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9504/
      figure1.ps          3/24/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      simdmf01.ps         3/24/95

Figure 2:
    parep2.lbl.gov:/export/home/u0/cedrdv/merrill/public_html/figs/
      fig86b.csh          10/9/95
      fig86b.ps           10/9/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      simdmf02.ps         10/9/95

Figure 3:
    parep2.lbl.gov:/h/merrill/selvin/parep/spatial/
      kmap1b.s            10/9/95
      kmap1b.ps           10/9/95
    parep2.lbl.gov:/h/merrill/selvin/parep/spatial/poisson
      poisson.f           10/7/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      simdmf03.ps         10/9/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/poisson/
      poisson.f           10/7/95

Figure 4:
    parep2.lbl.gov:/h/merrill/selvin/parep/spatial/
      knear2b.s           10/9/95
      knear2b.ps          10/9/95
    cedr.lbl.gov:/CEDRCD/data1/merrill/docs/parep/cdc9510/
      simdmf04.ps         10/9/95
```

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720