Incorporating data on residential history for disease mapping

C. JEFFERY,[a,c] A. OZONOFF,[a,b] , J. MANJOURIDES,[a] AND M. PAGANO,[a]
[a] Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115,USA
[b] Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118,USA
[c] International Health Group, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, United Kingdom

HARVARD School of Public Health

BOSTON UNIVERSITY SCHOOL of Public Health

LSTM LIVERPOOL SCHOOL OF TROPICAL MEDICINE

I. Motivation

Background

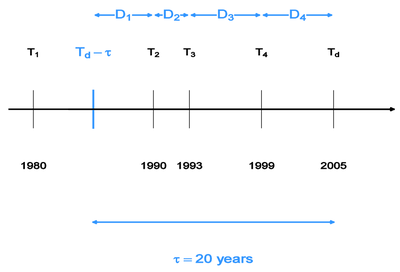
- Based on case location, disease mapping estimates a risk of disease across a geographic region
- Location at diagnosis does not necessarily correspond to location at exposure
- Some disease have long latency periods (e.g. leukemia, mesothelioma)
- Time, duration and location of exposure are unknown
- Residential history has already been incorporated in tests for cluster detection [1,3]

Questions

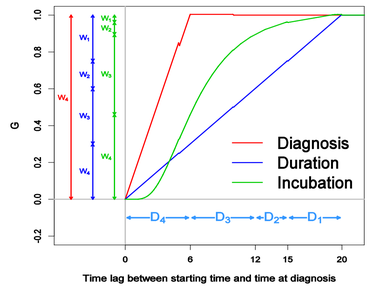
- How can residential history be incorporated in disease mapping?
- Can location at exposure be identified more accurately?

II. Methods

a. Example of data (4 locations)



b. Creating Weights: $W_k = G(D_k)$



c. Adapt distance-based mapping (DBM) [2]

- **Disease mapping**: compares an observed CDF F to an expected F_0 across a 2-dimensional study region
- **DBM** consists of four steps:
  1. Project the data to one dimension: Observed distribution of distances to one chosen fixed point (F_i)
  2. Compare the observed distribution to that expected under the null (F_{0i})
  3. Repeat 1 and 2 for a selection of fixed points (i =1,...,N)
  4. Average the measure across projections to compute a risk-like score at each point in the region
- **DBM adapted to residential history**: replace F_i and F_{0i} by averaged sums $\Sigma W_k F_{ik}$ and $\Sigma W_{0k} F_{0ik}$ respectively

d. Simulations

- Expected spatial distribution (F_0) : Uniform in unit square
- Observed spatial distribution (F): Increased risk in small circle
- Strength: q% cases have one location in circle (Multinomial, incubation weights)

e. Evaluation

- Estimate DBM scores across region
- Resolution = 50x50 grid points
- Dichotomize scores according to threshold *

(*) Threshold: we draw 100 sample of size 100 including a randomly located cluster of 10 points. Cluster radius is uniformly selected from (0.05,0.3). For each sample and a range of threholds, we select the one that minimizes the distance between the points (1,0) and (sensitivity,1-specificity). The median threshold across all 100 simulations is then selected for all remaining simulations.

Number of grid points	With a high score	With a low score
In cluster region	a	b
Not in cluster region	c	d

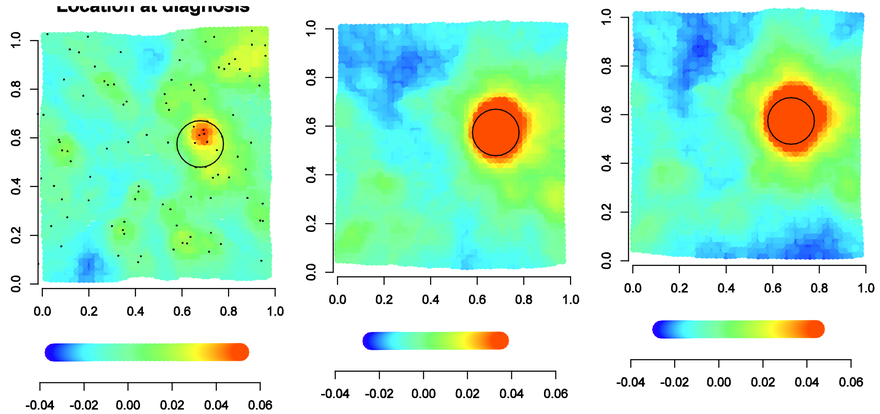
Sensitivity = a/(a+b) Specificity = d/(c+d)

Thanks and acknowledgements

Thanks to JSM 2010, conference organizers and participants
Research partially supported by:
• National Institutes of Health grant R01 EB0006195
• Centers for Disease Control and Prevention grant R01 PH000021-01

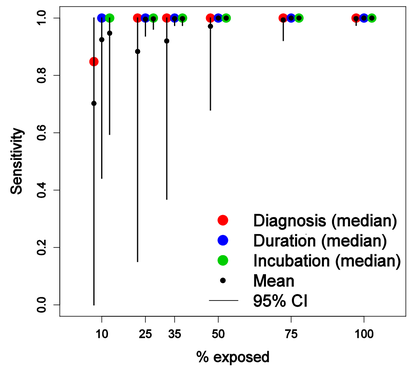
III. Results

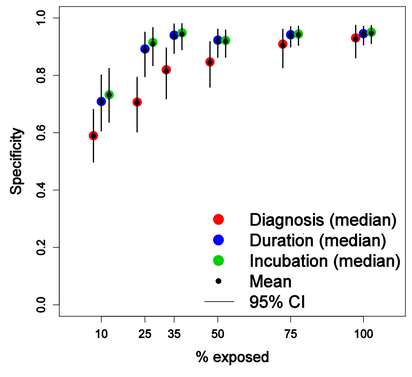
a. Illustration with 1 simulation (q=50%)



- The higher risk circle is identified more accurately by mapping using duration or incubation weights, rather than by mapping using only location at diagnosis
- The color cutoffs are determined by resampling from the reference population F_0

b. Evaluation of 1000 simulations





- Mapping using duration weights rather than only location at diagnosis improves sensitivity and specificity
- Mapping using incubation rather than duration weights improves sensitivity and specificity mostly when less than 50% cases are exposed
- Sensitivity and specificity tend to increase as the percentage of exposed cases increases

IV. Conclusion

- Disease mapping can incorporate residential history of cases by using a weighting scheme
- The accuracy at locating an increased risk improves by mapping with duration or incubation weights rather than mapping with location at diagnosis only
- There are other choices for the function G:
  - Step function
  - Weight all locations of a case equally
  - Include (time varying) covariates
- A similar method can be developed when cases are available with multiple daily locations (home/work/school) along with the proportion of the day spent at each location
- In future work, we can relax some limitations in the methods (missing spatial information) and simulations (non-uniform population, atemporal dichotomized risk)

References

1. Jacquez GM, Kaufmann A , Meliker , Goovaerts P , AvRuskin G , and Nriagu J. “Global, local and focused geographic clustering for case-control data with residential histories.” *Environmental Health: A Global Access Science Source* (BioMed Central Ltd) 4, no. 1 (2005): 4.
2. Jeffery C, Ozonoff A, White LF, and Pagano M. “Locating spatial clusters in a surveillance setting.” *Submitted.*
3. Manjourides J, and Pagano M. “Improving the power of chronic disease surveillance by incorporating residential history.” *Submitted.*