THE ROLE OF GEOSTATISTICS IN ENVIRONMENTAL EPIDEMIOLOGY

Pierre Goovaerts, PhD
Chief Scientist, BioMedware

International Association of Mathematical Geology
2013 Distinguished Lecturer
What is IAMG?
The mission of the IAMG is to promote, worldwide, the advancement of mathematics, statistics and informatics in the Geosciences.

Current & Upcoming Events

CoDaWork 2013
5th International Workshop on Compositional Data Analysis
Vorau, Austria, 3-7 June 2013

IAMG 2013 Annual Conference
in Madrid, Spain, 2-6 September 2013

IAMG 2014 Annual Conference
Jawaharlal Nehru University of New Delhi, India, 17-20 October 2014

See Meetings for other additional events!

New: The December 2012 Newsletter #85
is available for downloading from the website
Low (1.3 Mb) or Medium (3.7Mb) or High Resolution (8.4 Mb)
IAMG

- Founded in 1967
- Three journals
  - Mathematical Geosciences
  - Computers & Geosciences
  - Natural Resources Research
- 500+ members
- Annual conferences (Madrid in 2013, New Delhi in 2014)
- Affordable membership & journal subscription fees
REQUIRED (if not already member): IAMG DUES (for membership benefits, see the last page of this form)

- Ordinary member – 1 year .......................................................... US$ 12.00
- Ordinary member – 4 years .......................................................... US$ 48.00
- Ordinary member – Lifetime ......................................................... US$ 144.00
- Student member (proof of enrollment required) .......................... US$ 6.00
- Institutional member (includes library subscriptions to all 3 journals) .................................................................................. US$ 3,500.00

OPTIONAL: IAMG Journals

- Mathematical Geosciences (print and electronic versions), regular .......................................................... US$ 45.00
- Mathematical Geosciences (print and electronic versions), student (proof of enrollment required) ........ US$ 22.50
- Computers & Geosciences (print and electronic versions), regular ........................................................ US$ 105.00
- Computers & Geosciences (print and electronic versions), student (proof of enrollment required) ........ US$ 52.50
- Computers & Geosciences (electronic version only), regular or student .................................................. US$ 52.50
- Natural Resources Research (print and electronic versions), regular or student ....................................... US$ 55.00

OPTIONAL: IAMG Monograph Series

- “Computers in Geology: 25 Years of Progress” edited by J.C. Davis and U.C. Herzfeld .................. US$ 125.50
- “Modern Spatiotemporal Geostatistics” by George Christakos ............................................................. US$ 82.50
- “Geostatistical Analysis of Compositional Data” by V. Pawlowsky-Glahn and R.A. Olea .................. US$ 82.50
- “Statistical Methods for Estimating Petroleum Resources” by P. J. Lee ............................................... US$ 93.75

OPTIONAL: IAMG Conference Proceedings

Note: Proceedings are freely downloadable to members at: http://www.iamgmembers.org/phpbboard/
- Proceedings IAMG 2001 – 2011, C&G Silver – DVD .......................................................... US$ 15.00
- Proceedings IAMG 2001 – 2011, C&G Silver – Pen Drive .................................................. US$ 20.00

http://www.iamg.org/
Scientific journals
Stat conferences
Consulting
Eng. conferences
Software dvpt
Short courses

Well my fellow anthroposophists, I'm sure my lecture has made you think.

I stopped listening after 5 minutes.
What a bore.
ZZZZ....
I need a wee wee.

"I program, therefore I am...."
Outline

- What is geostatistics?
- Main steps of environmental geostatistical analysis
- Health geostatistical analysis
  - Spatial support
  - Small number problem
  - Modifiable areal unit problem
  - Combining areal and point data
  - Combining two sets of areal data
- Multi-scale joinpoint regression analysis
- Conclusions
Geostatistics

Geostatistics = Theory of regionalized variables
⇒ Statistical tools for analyzing space/time information

Application fields:

- Mining, petroleum, geophysics, geochemistry
- Soil science, forestry, agriculture (precision farming)
- Hydrology, oceanography, meteorology
- Remote sensing, environmental sciences (GIS, air, water & soil pollution)
- Health sciences (cancer data, disease)
Geostatistical Analysis

Data, ancillary information

↓

Model of Spatial Variability

↓

Spatial Prediction

↓

Model of Uncertainty

↓

Decision-making

→ cleaning

additional sampling
Geostatistical Analysis

Data, ancillary information

Model of Spatial Variability

Spatial Prediction

Model of Uncertainty

Decision-making

additional sampling
Variety of Analyzed Data

- Data measured in 2 or 3 dimensions (contaminated sediments)
- Space-time information
- Multiple layers of secondary information (DEM, geological map, satellite imagery)
Geostatistical Analysis

Data, ancillary information

Model of Spatial Variability

Spatial Prediction

Model of Uncertainty

Decision-making

cleaning

additional sampling
Description of Spatial Patterns

Semivariogram: measure of average dissimilarity between observations as a function of separation distance + direction.

\[ \gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_\alpha) - z(u_\alpha + h)]^2 \]

Nugget effect: discontinuity at the origin

Range: distance of correlation

Sill: plateau of bounded semivariograms
Geostatistical Analysis

Data, ancillary information
↓
Model of Spatial Variability
↓
Spatial Prediction
↓
Model of Uncertainty
↓
Decision-making

additional sampling

cleaning
**Kriging:** Generalized least-squares interpolation

\[ z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha (\mathbf{u}) z(\mathbf{u}_\alpha) \]

Multiple variants function:
- prediction support
- information available
- objectives (filtering...)
Geostatistical Analysis

Data, ancillary information

↓

Model of Spatial Variability

↓

Soil [Cd]  Spatial Prediction  Prediction variance
Geostatistical Analysis

Data, ancillary information

Model of Spatial Variability

Soil [Cd] Spatial Prediction Prediction variance
Change of Support

Point support

Kriging

Up-scaling

Side-scaling

Down-scaling

Block support

Block support
Geostatistical Analysis

Data, ancillary information

Soil [Cd]

Rock type map

N

1 km

Argovian
Kimmeridgian
Portlandian
Quaternary
Selandian
Heavy Metals in Soil

Field data

Soil map

Kriging

Kriging estimate

Kriging variance

BioMedvec
Geospatial Research and Applications
Geostatistical Analysis

Data, ancillary information

Model of Spatial Variability

Spatial Prediction

Model of Uncertainty

Decision-making

additional sampling cleaning
Modeling of Uncertainty

Spatial Uncertainty

Realization # 1

Realization # 2

Realization # 3

Local Uncertainty

Realization variograms

Indicator approach

Cd concentrations / mg kg\(^{-1}\)

[Graphs and maps showing spatial and local uncertainty with variograms and probability distributions.]
Modeling of Uncertainty

Spatial Uncertainty

Realization # 1

Realization # 2

Realization # 3

Local Uncertainty

Indicator approach

Probability of contamination

Realization variograms

Histogram

Cd concentrations / mg kg\(^{-1}\)
Kriging versus Simulation
Geostatistical Analysis

Data, ancillary information

Model of Spatial Variability

Spatial Prediction

Model of Uncertainty

Decision-making

Hazardous safe

BioMedware

Geospatial Research and Software

cleaning

additional sampling
Delineation of Contaminated Areas
Delineation of Contaminated Areas
Delineation of Contaminated Areas

B. Depth = 1.0 ft

- No pollution
- Pb ≥ 80
- As ≥ 9.3
- As ≥ 9.3 & Pb ≥ 80
Sampling Design
Medical Geography

Study of spatial patterns of disease incidence and mortality and the identification of potential “causes” of disease, such as environmental exposure or socio-demographic factors.

1. Punctual data
   Geo-referenced cases/controls, cancer stage
   **Information:** spatial coordinates

2. Spatially aggregated
   (ZIP codes, county, state)
   **Information:** mortality or incidence rate
Types of Spatial Data

1. Geostatistical Data

2. Lattice Data

3. Point Patterns
Health Geostatistical Analysis

Health Data

↓

Model of Spatial Variability

↓

Spatial Prediction (mapping, noise-filtering)

↓

Model of Uncertainty

↓

Decision-making

Cluster detection

Boundary detection

Correlation analysis
Health Data

Aggregated data

Individual-level data

Incidence of late-stage diagnosis (Breast cancer)
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data

Area-to-Area kriging
Any measurement & prediction relate to a non-zero, finite sample volume ($v$) which is too often assimilated to a point ($u$) in practice.

Geostatistical analysis is performed using the centroid geography.

Shape and size of counties is ignored in the analysis (assumption that the entire county population lives at a single location).
Spatial Support in Geostatistics

Any measurement & prediction relate to a non-zero, finite sample volume ($v$) which is too often assimilated to a point ($u$) in practice.

Geostatistical analysis is performed using the centroid geography.

Shape and size of counties is ignored in the analysis (assumption that the entire county population lives at a single location).
Spatial Support in Geostatistics

Any measurement & prediction relate to a non-zero, finite sample volume \((v)\) which is too often assimilated to a point \((u)\) in practice.

Theory for integrating various supports in kriging is described in Mining Geostatistics (1978)
Point Kriging: a Recall

\[ z^*(u_0) = \sum_{\alpha=1}^{5} \lambda_\alpha z(u_\alpha) + (1 - \sum_{\alpha=1}^{5} \lambda_\alpha) m \]

Data-to-data covariance

Data-to-unknown covariance

\[
\begin{bmatrix}
\rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} \\
\rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} & \rho_{25} \\
\rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} & \rho_{35} \\
\rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} & \rho_{45} \\
\rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & \rho_{55}
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\lambda_4 \\
\lambda_5
\end{bmatrix}
=
\begin{bmatrix}
\rho_{10} \\
\rho_{20} \\
\rho_{30} \\
\rho_{40} \\
\rho_{50}
\end{bmatrix}
\]
Matrix Notation of SK System

Kriging allows one to account for:

1. Pattern of spatial variability
   Zero weight beyond beyond range

2. Proximity of data to $u_0$

3. Data configuration
   Redundancy of $u_2$ and $u_3$
Another Notation for Kriging System

System of $n(u_0)$ equations with $n(u_0)$ unknowns
Area-To-Area Kriging

\[ z(B) = \sum_{k=1}^{K} \lambda_k z(v_k) \]

- Data = areal values (e.g. ZIP codes) and prediction performed over another set of possibly non-nested geographical units (e.g. census tracts)
- Prediction variance is computed for each block.
Area-To-Area Kriging

\[ \sum_{k'=1}^{K} \lambda_k \overline{C}(v_k, v_{k'}) = \overline{C}(v_k, B) \quad k = 1, \ldots, K \]

**Numerical approximation of** \( \overline{C}(v_i, v_j) \)

\[ \overline{C}(v_i, v_j) = \frac{1}{P_i P_j} \sum_{s=1}^{P_i} \sum_{s'=1}^{P_j} C(u_s, u_{s'}) \]
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data
   Area-to-Area kriging

2. Instability of rates due to small populations
   Leelanau peninsula
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data
Area-to-Area kriging

2. Instability of rates due to small populations
Poisson & binomial kriging (kriging with non-systematic measurement errors)
A Probabilistic Model

**d(ν_α):** Number of recorded mortality cases in entity ν_α

outcome of a Poisson count variable

[D(ν_α)| R(ν_α) ~ Poisson {n(ν_α)×R(ν_α)}] \( \alpha=1,\ldots,N \)

Expected number of counts

**n(ν_α):** Population at risk

**R(ν_α):** Local risk
Estimation of Risk Values

Poisson kriging (Monestiez et al., 2006)

\[ \tilde{r}(v_\alpha) = \sum_{i=1}^{s(v_\alpha)} \lambda_i(v_\alpha) z(v_i) \]

Observed frequencies \( \frac{d(v_i)}{n(v_i)} \)

Error variance term

\[ \sum_{j=1}^{K} \lambda_j(v_\alpha) \left[ \bar{C}_R(v_i, v_j) + \delta_{ij} \frac{m^*}{n(v_i)} \right] + \mu(v_\alpha) = \bar{C}_R(v_i, v_\alpha) \quad i = 1, \ldots, K \]

\[ \sum_{j=1}^{K} \lambda_j(v_\alpha) = 1 \]
Binomial Kriging

\[ \hat{r}_{BK}(v_\alpha) = \sum_{i=1}^{K} \lambda_i(v_\alpha) z(v_i) \]

\[ \sum_{j=1}^{K} \lambda_j(v_\alpha) \left[ \overline{C}_R(v_i, v_j) + \delta_{ij} \frac{a}{n(v_i)} \right] + \mu(v_\alpha) = \overline{C}_R(v_i, v_\alpha) \quad i = 1, ..., K \]

\[ \sum_{j=1}^{K} \lambda_j(v_\alpha) = 1. \]

\[ a = m^*(1-m^*) - \overline{C}_I(v_i, v_i) \]
Breast Cancer Mortality
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data
   Area-to-Area kriging
2. Instability of rates due to small populations
   Poisson & binomial kriging
3. No risk data for variogram estimation
   Population-weighted estimator
4. Point-support variogram unavailable
   Iterative deconvolution procedure
Estimation of Covariance of the Risk

Variogram of unknown risk derived from variogram of observed frequencies:

\[
\hat{\gamma}_R(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} \frac{\sum_{\alpha} \{ n(v_\alpha)n(v_\alpha + h) \}}{n(v_\alpha) + n(v_\alpha + h)} \left\{ \frac{n(v_\alpha)n(v_\alpha + h)}{n(v_\alpha) + n(v_\alpha + h)} \left[ z(v_\alpha) - z(v_\alpha + h) \right]^2 - \bar{z} \right\}
\]
Variogram and Change of Support

Pixel = 1×1
Variogram and Change of Support

Pixel = 1×1

Pixel = 2×2
Variogram and Change of Support

\[ \gamma(h) = \gamma(h) - \bar{\gamma}(v, v) \]
How to Derive point-support $\gamma(h)$?

Only areal data are available $\Rightarrow$ estimation of regularized $\gamma_v(h)$

Followed by Deconvolution

$$
\gamma_v(h) = \gamma(h) - \bar{\gamma}(v,v)
$$

Classical deconvolution formula is based on the assumption that all the blocks/areas have the same size and shape: $\bar{\gamma}(v,v) = Cst$
Deconvolution for Irregular Areas

\[ \gamma_v(h) = \gamma(v, v_h) - \gamma_h(v, v) \]

Both area-to-area variograms are a function of distance \( h \) since small counties tend to be paired for shorter separation distances.

\[ \bar{\gamma}(v, v_h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} \bar{\gamma}(v_{\alpha}, v_{\alpha+h}) \]

\[ \bar{\gamma}_h(v, v) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[ \bar{\gamma}(v_{\alpha}, v_{\alpha}) + \bar{\gamma}(v_{\alpha+h}, v_{\alpha+h}) \right] \]
Iterative Deconvolution Procedure

1. Start with candidate point-support model

2. Regularize model using formula and compare it to the model inferred from data

3. Based on differences, rescale the optimal model found so far ⇒ new candidate for next iteration

\[ \text{Risk (areal data)} \quad \text{Deconvoluted model} \quad \text{Regularized model} \]
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data
   Area-to-Area kriging

2. Instability of rates due to small populations
   Poisson & binomial kriging (kriging with non-
   systematic measurement errors)

3. No risk data for variogram estimation
   Population-weighted estimator

4. Point-support variogram unavailable
   Iterative deconvolution procedure

5. Arbitrary spatial support for data (MAUP)
Challenges

• Modifiable Areal Unit Problem
Challenges & Solutions

1. Irregular spatial (non-point!) supports for data
   Area-to-Area kriging

2. Instability of rates due to small populations
   Poisson & binomial kriging

5. Arbitrary spatial support for data (MAUP)
   Area-to-Point kriging

Choropleth maps => isopleth maps
Area-To-Point Kriging

\[ \hat{r}(u) = \sum_{i=1}^{s(u)} \lambda_i(v_\alpha)z(v_i) \]

- Data = areal values (county-level rates) and prediction performed at point location.
- **Consistency:** average of point estimates within each block returns the areal value as long as same K areal data \( V_k \) are used for all points within the block.
- Prediction variance is computed for each grid node.
Area-To-Point Poisson Kriging

\[ \sum_{j=1}^{K} \lambda_j(u_\alpha) \left[ \overline{C}_R(v_i, v_j) + \delta_{ij} \frac{m^*}{n(v_i)} \right] + \mu(v_\alpha) = \overline{C}_R(v_i, u_\alpha) \quad i = 1, \ldots, K \]

\[ \sum_{j=1}^{K} \lambda_j(v_\alpha) = 1 \]

Area-to-Area covariance

Area-to-Point covariance

Numerical approximation of \( \overline{C}_R(v_i, u_\alpha) \)

\[ \overline{C}_R(v_k, u) = \frac{1}{P_i} \sum_{s=1}^{P_i} C_R(u_s, u) \]
Health Data

Aggregated data

Individual-level data

Incidence of late-stage diagnosis (Breast cancer)
Mapping Health Outcomes

Aggregated data

Late stage diagnosis

Individual-level data

Late stage diagnosis

Binomial kriging

Indicator kriging
How can Geostatistics help us?

Indicator approach can be applied to mapping frequency of late-stage Breast cancer diagnosis

\[ i(u_\alpha) = \begin{cases} 
1 & \text{if late stage diagnosis at } u_\alpha \\
0 & \text{otherwise} 
\end{cases} \]

\[ \hat{I}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [i(u_\alpha) - i(u_\alpha + h)]^2 \]

\[ i^*(u) = \sum_{\alpha=1}^{n} \lambda_\alpha i(u_\alpha) \]
Challenges

- Wide range of distances sampled

- 2-step modeling procedure
- Use of rescaled horizontal axis

Late detection cases do not occur randomly in space, yet individual-level factors such as age or family history generate a large variability over very short distances (1st range = 48 m).
Area-And-Point kriging

\[ r^*(u) = \sum_{\alpha=1}^{n} \lambda_\alpha i(u_\alpha) + \sum_{k=1}^{K} \beta_k z(v_k) \]

Combination of both point and areal data ⇒ solving one kriging system

\[ \sum_{j=1}^{(n+K)} \lambda_j \overline{C}(x_i, x_j) = \overline{C}(x_i, u) \quad i = 1, \ldots, (n + K) \]

Area-to-Area covariance
Area-to-Point covariance
Point-to-Point covariance

\[ x_i = v_i \quad x_j = v_j \]
\[ x_i = v_i \quad x_j = u_j \]
\[ x_i = u_i \quad x_j = u_j \]
Kriging with Area-based & Individual-level Data

Ordinary IK

AAP Kriging

Census rates
Area(s)-To-Point Binomial Kriging

\[
\hat{r}(u) = \lambda_{\alpha'} z(\nu_{\alpha'}) + \sum_{\alpha=1}^{K'} \lambda_{\alpha} z(\nu_{\alpha}) + \lambda_{\beta'} y(\nu_{\beta'})
\]

Kernel primary data secondary data
ZIP-code ZIP-code Census

primary data are the most reliable.
Late-stage diagnosis data are available at:

- **Census-tract** level
  - Individual-level residence coordinates were randomized within each tract for **confidentiality reasons**.

- **Zip code** level
  - Include all records, including the residences that **failed to geocode**.
Results: Kriging Estimates & Variance

Zip code-level % late-stage diagnosis

Census tract-level % late-stage diagnosis

ATP binomial kriging (ZIP code and tract data)

ATP binomial kriging variance (ZIP code and tract)

\[ \sigma_K^2(\mathbf{u}) = C(0) - \sum_{i=1}^{K+2} \lambda_i \overline{C}(v_i, \mathbf{u}) - \mu(\mathbf{u}) \]
Results: Kriging Estimates

Zip code-level % late-stage diagnosis

Census tract-level % late-stage diagnosis

ATP binomial kriging (ZIP code and tract data)

County-level % late-stage diagnosis
Results: Kriging Estimates

Zip code-level % late-stage diagnosis

ATP binomial kriging (ZIP code and tract data)

Census tract-level % late-stage diagnosis

County-level % late-stage diagnosis
Geographical, Temporal and Racial Disparities in Late-stage Prostate Cancer Incidence across Florida: A Multiscale Joinpoint Regression Analysis

Pierre Goovaerts, PhD
Chief Scientist, BioMedware
Outline

- Setting the problem
  - Prostate cancer: temporal trend in Florida
  - Geographical & Racial disparities
- Quantitative analysis
  - Joinpoint regression
  - Disparity statistics
- Multi-scale Analysis
  - State-level analysis
  - Metropolitan vs non-metropolitan areas
  - County-level analysis

Conclusions
1. **Prostate cancer** is the most frequently diagnosed non-skin cancer and the 2nd leading cause of male cancer-related death in the US.

2. Prostate cancer mortality and late–stage diagnosis started **declining after 1991**, which is likely due to early detection (PSA screening).
3. Existence of substantial variation among counties for the timing and magnitude of decline, as well as average % late-stage.
4. Existence of substantial geographical & racial disparities in temporal trends, as well as average % late-stage.
**Aim:** Model each time series using a few continuous linear segments.

**Joinpoint Regression**

- $\beta_1 = -0.004975$
- $\beta_2 = -0.1019$
- $\beta_3 = 0.01794$

$\tau_1 = 1989$, $\tau_2 = 2000$

**Joinpoints** represent the timing for a statistically significant change in rate trend.

**APC (Annual Percent Change):** slope of segment
Joinpoint Regression

1. **Estimation** of number/value of joinpoints & APC through an iterative procedure that tests whether models of increasing complexity yield significantly better goodness-of-fit than simpler models.

2. The models may be linear on log(response) → compute **annual rate percentage change**.

3. Two segmented linear regressions (e.g. time trends for two ethnic groups or two counties) can be compared and the parallelisms or identity of the two mean functions can be tested.
Comparison of Time-series

**Approach 1:** Test whether the two time series are parallel or not
Comparison of Time-series

* denotes parameters significantly different from zero at $\alpha=0.05$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th></th>
<th>White males</th>
<th></th>
<th>Black males</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>CI$_{0.025}$</td>
<td>CI$_{0.975}$</td>
<td>Estimate</td>
<td>CI$_{0.025}$</td>
</tr>
<tr>
<td>Joinpoint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1981, $\tau_1$]</td>
<td>-0.5</td>
<td>-1.6</td>
<td>0.6</td>
<td>0.5</td>
<td>-3.0</td>
</tr>
<tr>
<td>[$\tau_1$, $\tau_2$]</td>
<td>-9.7*</td>
<td>-10.4</td>
<td>-9.0</td>
<td>-10.2*</td>
<td>-12.2</td>
</tr>
<tr>
<td>[$\tau_2$, 2007]</td>
<td>1.8</td>
<td>-0.2</td>
<td>3.8</td>
<td>-0.2</td>
<td>-4.3</td>
</tr>
<tr>
<td>AAPC</td>
<td>1981-2007</td>
<td>-3.9*</td>
<td>-4.5</td>
<td>-3.3</td>
<td>-4.4*</td>
</tr>
</tbody>
</table>
Comparison of Time-series

**Approach 1**: Global Test of whether the two time-series are parallel or not.

**Approach 2**: Compute the number of years where the confidence intervals of the APC parameters do not overlap.

\[ B_{rr'} = \sum_{t=1}^{T} I((\text{CI}(r; t); \text{CI}(r'; t))) \]  

where the indicator function \( I(.) = 1 \) if the following condition on the upper bounds \((U)\) and lower bounds \((L)\) of the two confidence intervals CI are met: \( U(r; t) < L(r'; t) \) or \( L(r; t) > U(r'; t) \). A large number indicates that rates of changes for the two races are consistently different over time. There is no statistical test associated with quantity \( B_{rr'} \), which is mainly descriptive.
## State-level Racial Disparities

<table>
<thead>
<tr>
<th>Parameter</th>
<th>White males</th>
<th></th>
<th>Black males</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>CI&lt;sub&gt;0.025&lt;/sub&gt;</td>
<td>CI&lt;sub&gt;0.975&lt;/sub&gt;</td>
<td>Estimate</td>
</tr>
<tr>
<td><strong>Joinpoint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>APC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1981, ( \tau_1 )]</td>
<td>-0.5</td>
<td>-1.6</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>([ \tau_1 , \tau_2 ])</td>
<td>-9.7*</td>
<td>-10.4</td>
<td>-9.0</td>
<td>-10.2*</td>
</tr>
<tr>
<td>([ \tau_2 , 2007 ])</td>
<td>1.8</td>
<td>-0.2</td>
<td>3.8</td>
<td>-0.2</td>
</tr>
<tr>
<td><strong>AAPC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981-2007</td>
<td>-3.9*</td>
<td>-4.5</td>
<td>-3.3</td>
<td>-4.4*</td>
</tr>
</tbody>
</table>

\[ \text{B}_{rr'} = 2 \text{ years (non-overlapping CIs in 1989 & 1999)} \]
Rural versus Urban Areas

Average Beale code

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Counties in metro areas of 1 million population or more</td>
</tr>
<tr>
<td>2</td>
<td>Counties in metro areas of 250,000 to 1 million population</td>
</tr>
<tr>
<td>3</td>
<td>Counties in metro areas of fewer than 250,000 population</td>
</tr>
</tbody>
</table>

Metropolitan counties

Non-metropolitan counties

1. Urban population of 20,000 or more, adjacent to a metro area
2. Urban population of 20,000 or more, not adjacent to a metro area
3. Urban population of 2,500 to 19,999, adjacent to a metro area
4. Urban population of 2,500 to 19,999, not adjacent to a metro area
5. Completely rural or less than 2,500 urban population, adjacent to a metro area
6. Completely rural or less than 2,500 urban population, not adjacent to a metro area
Rural versus Urban Areas

provider (family and internal medicine, urology)
Rural versus Urban Areas

Metropolitan counties

\( B_{rr'} = 1 \text{ years} \)

Non-metropolitan counties

\( B_{rr'} = 12 \text{ years} \)
County-level Analysis

Prostate cancer late-stage diagnosis (1981-2007)

Map of Florida showing percentage of late-stage diagnoses by county.

- Percentage late-stage diagnoses for specific counties over time.
County-level Analysis

A. Average APC (WM)
   - APC:
     - Red: 0.0 to 1.50
     - Dark Red: -1.5 to 0.0
     - Orange: -3.0 to -1.5
     - Yellow: -4.5 to -3.0
     - Light Yellow: -6.0 to -4.5
     - Grey: -7.5 to -6.0
     - Blue: <=-7.5

B. Onset year of decline (WM)
   - Year:
     - Red: 1994 to 1997
     - Orange: 1991 to 1994
     - Light Orange: 1988 to 1991
     - Grey: 1985 to 1988
     - Blue: 1982 to 1985

C. Average APC (BM)
   - APC:
     - Red: 0.0 to 1.50
     - Dark Red: -1.5 to 0.0
     - Orange: -3.0 to -1.5
     - Yellow: -4.5 to -3.0
     - Light Yellow: -6.0 to -4.5
     - Grey: -7.5 to -6.0
     - Blue: <=-7.5

D. Onset year of decline (BM)
   - Year:
     - Red: 1994 to 1997
     - Orange: 1991 to 1994
     - Light Orange: 1988 to 1991
     - Grey: 1985 to 1988
     - Blue: 1982 to 1985
County-level Analysis

- Miami-Dade
- Duval
- Hillsborough
- Orange
- Escambia
- Alachua
- $3 < \text{Beale} \leq 6$
- $\text{Beale} > 6$
County-level Analysis

- Miami-Dade
- Duval
- Hillsborough
- Orange
- Escambia
- Alachua
- 3 < Beale ≤ 6
- Beale > 6

Brr = 2

Miami

Brr = 6

Alachua

Brr = 16

Escambia

Brr = 16

Orange

Brr = 8

Hillsborough

Brr = 9

Duval

Brr = 7

Beale 3-6

Brr = 25

Beale > 6

BioMedware
Geospatial Research and Software
Impact of Scale

Magnitude of racial disparities at smaller spatial scale where SES & access to care is more uniform?
Conclusions

1. Both ethnic groups experienced a 50% decline in the state-average proportion of late-stage diagnosis. This drop started in the early 1990s when PSA became widely available.

2. Analysis at the metropolitan and non-metropolitan levels revealed that the frequency of late-stage diagnosis increased recently in urban areas, and this trend was significant for white males.

3. Annual rate of decrease in late-stage diagnosis and the onset years for significant declines varied greatly among counties and racial groups.

4. Spatial extent of racial disparities reached a peak in 1990 because of an early decline in frequency of late-stage diagnosis observed for black males.

http://www.iamg.org/