Technical Feasibility of Establishing a Proactive Cancer Cluster Surveillance System Using the SPAS Macro

Jesse J. Plascak  The Ohio State University
Overview

1. Barriers to conducting state-wide Non-communicable disease cluster (NCC) investigations
2. SaTScan’s™ robustness
3. SaTScan™-Python-ArcGIS®, via SAS® (SPAS) macro
4. SPAS example within SAS®
5. Implementation of SPAS is efficient
NCC Investigative Needs¹

1. “Provide technical assistance and training”
2. “Increase staffing and budgets”
3. “Develop more robust statistical methods for assessing potential clusters”
4. “Improve tools, such as [GIS], to effectively support NCC analysis and response”

SPAS Purpose

- To process a cancer registry dataset for spatial clustering analysis and subsequent geoprocessing of clustering results in the most efficient yet technically comprehensible way possible
SaTScan’s™ Robustness

- SaTScan™ space-time cluster software
- Location *and* significance of clusters
- Improvements from v1.0 – v9.1
  - Elliptical scan
  - Gumbel distribution
  - Multiple-run adjustment
  - Covariate adjustment
  - Clusters within clusters
  - Straightforward output
SaTScan’s™ Robustness

- Demonstrable statistical power
- Computationally efficient
- Flexibility of options
- User-friendly
- Free
SaTScan™-Python-ArcGIS® via SAS® (SPAS)

- Based on previous SAS® macro\(^2\)
- Five primary functions:
  1. Process input dataset
  2. Create SaTScan™ files
  3. Create Python geoprocessing file
  4. Run SaTScan™
  5. Geoprocess SaTScan™ results via Python

SPAS General Characteristics

- SAS® for Windows
- Compatible with SaTScan™ v9.1
- Requires installation of ArcGIS® (Python)
- Current SPAS version supports:
  - Spatial analyses only
  - Poisson model (aggregate counts/population)
  - Census tract level
Hypothetical Dataset

- Random cases by CT – Franklin CO, Ohio cluster
SPAS Specifics: Process Input Dataset

- Aggregates counts by state-county-census tract
- Joins to geography/population dataset (FID)
- Error dataset where cases > population
- Requires 7 macro variable inputs
SPAS Specifics: Process Input Dataset

- Appearance in SAS®

4) !REQUIRED - MODIFY ONCE! datadir = The directory storing the initial sas dataset to be scanned. Also where all files created from this program will be stored */

```
datadir=E:\NAACCR_2011\naaccr_pres_analysis\computation\sas_macro_depot,
/*
```

5) !REQUIRED! registry_dataset_name = The name of the sas dataset file containing the location-related event variables to run the SPAS macro */

```
registry_dataset_name=registry,
/*
```

6) !REQUIRED! clustering_dataset_name= A name to be used in all output file names */

```
clustering_dataset_name=home,
```
SPAS Specifics: SaTScan™ File Creation

- Two different SAS® macros
  1. .CAS, .POP, and .GEO files
SPAS Specifics: SaTScan™ File Creation

- Two different SAS® macros
  1. .CAS, .POP, and .GEO files

```
home_coord - Notepad

File  Edit  Format  View  Help
39001990100 38.957143596 -83.35253236
39001990200 38.982772599 -83.54938117
39001990300 38.84061958 -83.58287059
39001990400 38.773752667 -83.53580583
39001990500 38.755960648 -83.35666064
39001990600 38.697324338 -83.58155019
39003010100 40.891634866 -83.89953907
39003010200 40.862773657 -84.00119197
39003010300 40.82415675 -84.1657104
39003010600 40.731607677 -84.31313644
39003010800 40.759661373 -84.19722495
39003010900 40.744941585 -84.15705211
39003011000 40.777335464 -84.1344947
39003011200 40.78116319 -84.08885564
39003011300 40.776014043 -84.03853263
```
Two different SAS® macros

1. .CAS, .POP, and .GEO files
2. .PRM SaTScan™ parameter file
   - All (60+) parameter specifications
   - Most macro variables require some input
   - Few will change during routine surveillance
SPAS Specifics: SaTScan™ File Creation

- Two different SAS® macros
  1. .CAS, .POP, and .GEO files
  2. .PRM SaTScan™ parameter file

```plaintext
[Input]
CaseFile=E:\NAACCR_2011\naaccr_pres_analysis\computation\sas_macro_depot\home_cases.cas
PopulationFile=E:\NAACCR_2011\naaccr_pres_analysis\computation\sas_macro_depot\home_pop.pop
CoordinatesFile=E:\NAACCR_2011\naaccr_pres_analysis\computation\sas_macro_depot\home_coord.geo
UseGridFile=n
GridFile=
PrecisionCaseTimes=0
CoordinateType=1
StartDate=2009/01/01
EndDate=2009/12/31

[Analysis]
AnalysisType=1
ModelType=0
ScanAreas=1
TimeAggregationUnits=1
TimeAggregationLength=1

[Output]
ResultsFile=E:\NAACCR_2011\naaccr_pres_analysis\computation\sas_macro_depot\home_results.txt
SaveSimLLR=ASCII=n
SaveSimLLRDBase=y
IncludeRelativeRisksCensusAreasASCII=n
IncludeRelativeRisksCensusAreasDBase=y
CensusAreasReportedClustersASCII=n
CensusAreasReportedClustersDBase=y
```
SPAS Specifics: Python File Creation

- Writes Python code for subsequent geoprocessing of SaTScan™ results

```python
import sys, os, arcgisscripting

gp = arcgisscripting.create()
gp.workspace = "E:/NAACCR_2011/naaccr_pres_analysis/computation/sas_macro Depot"

in_table=r"E:/NAACCR_2011/naaccr_pres_analysis/computation/sas_macro Depot/home_results_col.cbf"
in_x="LONGITUDE"
in_y="LATITUDE"
out_layer="home_clusters.lyr"
sref=r"C:/Program Files (x86)/ArcGIS/Desktop10.0/Coordinate Systems/Geographic Coordinate Systems/North America/NAD 1983.prj"
gp.MakeXYEventLayer(in_table, in_x, in_y, out_layer, sref)
gp.SaveToLayerFile(out_layer, "E:/NAACCR_2011/naaccr_pres_analysis/computation/sas_macro Depot/home_clusters.lyr")
```
SPAS Specifics: Run SaTScan

- Using files created in preceding macros
- Calls batch SaTScan™ using SAS’s® X Command

X "cd C:\Program Files\Satscan";
X "SaTScanBatch64.exe E:\NAACCR_2011\naacccr_pres_analysis\computation\sas_macro_depot\example.prm";
SPAS Specifics: Run SaTScan

- Using files created in preceding macros
- Calls batch SaTScan™ using SAS’s® X Command
SPAS Specifics: Geoprocessing with Python

- Geoprocesses SaTScan™ results file
  1. Creates/saves layer of cluster centroids
  2. Creates/saves shapefile of cluster centroids
  3. Creates/saves shapefile of cluster radius buffer
  4. Projects shapefile (USA contiguous conic)
Hypothetical Dataset Results

- X Command opens pre-saved ArcMap™ base map and SaTScan™ results summary text file
- Newly created shapefiles can be added to ArcMap™ base map
Hypothetical Dataset Results
Hypothetical Dataset Results
Hypothetical Dataset Results
SPAS is Efficient

- Single program performs multiple tasks in sequence
- No physical interfacing with SaTScan™, Python or ArcGIS®
- Few macro variable changes beyond initial inputs
- Flexibility of GIS-compatible file creation
- SPAS can be expanded for further capability
SPAS Conclusion

SPAS accomplishes:

1. Processing of registry dataset for areal-based spatial cluster analysis
2. Utilization of the robust spatial clustering software – SaTScan™
3. Streamlining of basic GIS computations of SaTScan™ results
4. Minimization of technical involvement
5. Minimization of personnel investment
Collaborators

- Ohio Department of Health
  - Holly L. Sobotka, M.S.; Robert W. Indian, M.S.

- The Ohio State University Comprehensive Cancer Center – James Cancer Hospital and Solove Research Institute
  - James L. Fisher, Ph.D.; Julie A. Stephens, M.S.; Electra D. Paskett, Ph.D.
Thank You

Jesse J. Plascak
The Ohio State University
plascak.2@buckeyemail.osu.edu