WHONET

Cluster Detection with SaTScan™

WHO Collaborating Centre for Surveillance of Antimicrobial Resistance

Boston, July 2022

WHONET Tutorial – Cluster Detection with SaTScan™

For many years, WHONET has had the ability to present temporal trends in organism frequencies and resistance proportions in the form of descriptive statistics and graphs. The user could then examine these trends and graphs individually in efforts to detect and characterize possible community or hospital outbreaks of microorganisms. WHONET was not able to highlight potential clusters in an automatic fashion in order to focus the attention of the software user on possible outbreaks or to provide statistical guidance as to whether observed trends were statistically significant.

To facilitate the early and broad detection of possible outbreaks, we have integrated a powerful freeware tool developed for purposes of cluster detection in public health data. SaTScan™, a trademark of Martin Kulldorf, was developed under the joint auspices of Martin Kulldorf, of the National Cancer Institute and of Farzad Mostarashi at the New York City Department of Health and Mental Hygiene. Dr. Kulldorf is an Associate Professor and Biostatistician at Harvard Medical School and Harvard Pilgrim Health Care, Department of Ambulatory Care and Prevention, Boston, USA.

> Kulldorf M. and Information Management Services, Inc. SaTScan™ v.7.0: Software for the spatial and time-scan statistics. http://www.satscan.org, 2006.

The software permits a number of algorithms for the detection of event clusters. Options include retrospective or prospective cluster detection; purely temporal, pure spatial, or space-time clusters; and flexible parameter selection for space and time variables. In this first version of an integrated WHONET-SaTScan package, WHONET is using a space-time permutation probability model, using Monte Carlo simulations. In collaboration with Dr. Kulldorf in a five-year NIH project entitled "Modeling Infectious Disease Agent Study" (MIDAS), we will test out a number of additional algorithms, models, and parameters, and these optimized routines will eventually be offered through the WHONET user interface.

In WHONET 5.4, SaTScan statistics have been integrated into two standard analysis features: 1. Isolate listing summaries and 2. Resistance profile summaries. Two new analysis options specific to SaTScan have also been added: 1. SaTScan – Cluster detection in Data Analysis and 2. SaTScan – Space-and-Time cluster detection in Quick Analysis.

Part 1.   SaTScan and Isolate listing summaries

Part 2.   SaTScan and Resistance profile summaries

Part 3.   SaTScan – Cluster detection (Data Analysis)

Part 4.   SaTScan – Space-and-Time cluster detection (Quick Analysis)

Part 1.   SaTScan and Isolate listing summaries

Open the WHO Test Hospital, and go to Data Analysis. Go to Analysis Type, and select Isolate listing and summary. At the bottom of the screen, click on the option "Include SaTScan alerts". By doing this, the SaTScan statistics will be integrated directly into the standard Isolate Listing Summary output.

Note: In the above example, the column variable has been set to "Specimen date" by "Day". In most cases, you should probably leave this option set to "Month" or "Week". But for purposes of this tutorial, we have set the option to "Day" because the sample dataset only includes one month of data.

To see the details of the SaTScan analysis, click on "SaTScan" to see the below screen.
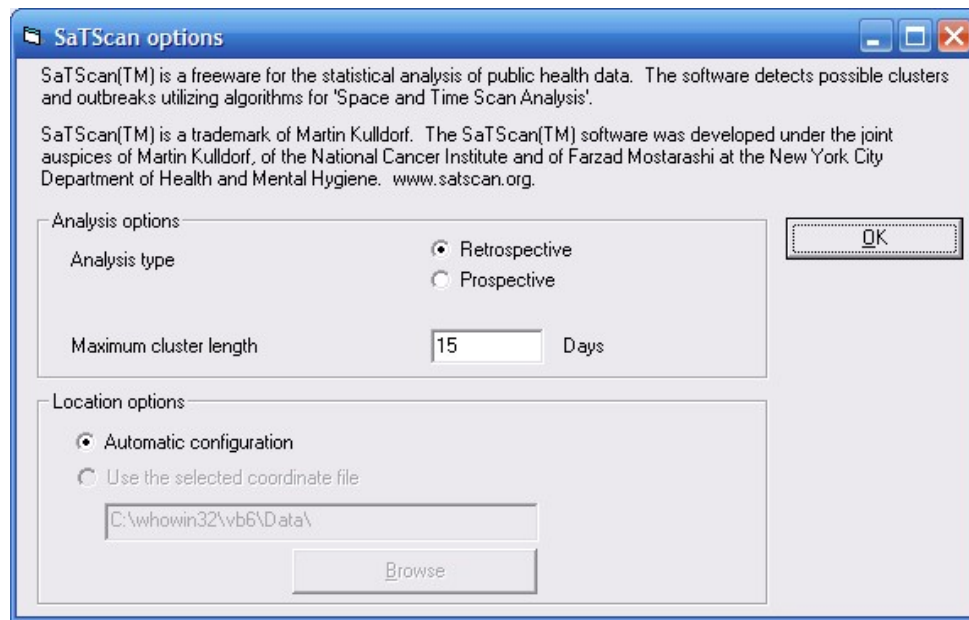
On this screen, you will see a small minority of all of the parameter options available in SaTScan. The use of SaTScan to study microbiological laboratory data and hospital infections data is very new, and will be a major focus of the 5-year NIH grant. As we gain more experience with the use of SaTScan for microbiology data, the number of options included in WHONET will increased and the suggested default parameter values will change to reflect optimal use of the program by microbiologists, infection control, and clinical staff.

In the above example, the user has access to two SaTScan variables.

Analysis type: *Retrospective vs. Prospective*. Retrospective analysis looks for infectious disease clusters at any time during the data analysis time period – both old clusters and new clusters. This option would be very useful in a retrospective look at available data for any possible outbreak. Prospective analysis only looks for clusters that potentially are still ongoing. This option is appropriate for the prospective detection of outbreaks in the new data.

Maximum cluster length: Indicate the number of days for the greatest length of a possible outbreak. At present, the default is set to 15 days. In other words, SaTScan will look for clusters that are anywhere from 1 day to 15 days in length. If an outbreak is longer than this period, SaTScan possibly may not detect it or possibly may detect only part of it. To find the greatest possible number of potential clusters, you can put a large value for the Maximum cluster length. A long period can slow SaTScan down considerably, so you can experiment to see what impact this value has on 1) the number of detected clusters and 2) the time period needed for the analysis.

Note: In SaTScan, the largest possible value of the Maximum cluster length is half the analysis time period. For example, if you have one year (365 days) of data to scan, the largest value that SaTScan will accept is 182 days. It is not a problem

if you enter a higher value into WHONET. WHONET will dynamically adjust the value before it sends the analysis parameters to SaTScan.

Location options: In WHONET so far, we are not yet taking advantage of SaTScan's ability to handle geographic data. For outpatients, such variables could include zip code or street address GPS latitude and longitude. For inpatients, it is possible describe to SaTScan the geographic (hospital physical location) or functional (clinical care service) relationship between patient care units. So at the present time, WHONET does an automatic configuration of patient locations, treating each ward as a separate "island" of care, unconnected to any other wards.

In the future, it will be possible for WHONET users to describe the ward structure of hospital units in a SaTScan "coordinate file" in a simple user interface. If a user does take the time to describe the geography of their hospital to WHONET, this option will permit the user to select the coordinated file created.

After you select the desired features, click "OK" to return to the main analysis screen.

Then click on "One per patient". Select the options "By patient", "First isolate only".

Note: It would be probably more useful for cluster detection to put the option "By time interval or resistance phenotype", for example with a 30 day window between isolates. This option does exist for %RIS calculation, but unfortunately, this option is not yet ready for other analyses, such as isolate listings and resistance profiles. This will be addressed in the near future.

Note: We do not yet know what criteria would be most effective in detecting real infectious disease outbreaks, and whether the criteria should differ for different organisms. SaTScan does an excellent job of finding statistically defined "clusters" of events, but the clinical interpretation of these clusters will depend on the experience and insight of infection control and clinical staff. With more testing on

real datasets, we will aim to provide tested and validated recommendations for setting the cluster detection parameters.

Finally, click "OK" and Begin Analysis. First you will see the standard isolate listing. (SaTScan makes no changes to the isolate listing. If you do not wish to see the listing, you should choose "Summary" as the Report format rather than "Both".)

When you click Begin Analysis, WHONET will read the data files, and prepare the files needed by SaTScan to operate. WHONET will then send this information to SaTScan. A small icon will appear at the bottom of the screen to tell you that SaTScan is running in the background. When SaTScan is finished, this icon will disappear automatically.

After SaTScan is finished analyzing the data, WHONET will read the SaTScan results, and integrate the statistics into the normal WHONET output, as in the below screen.
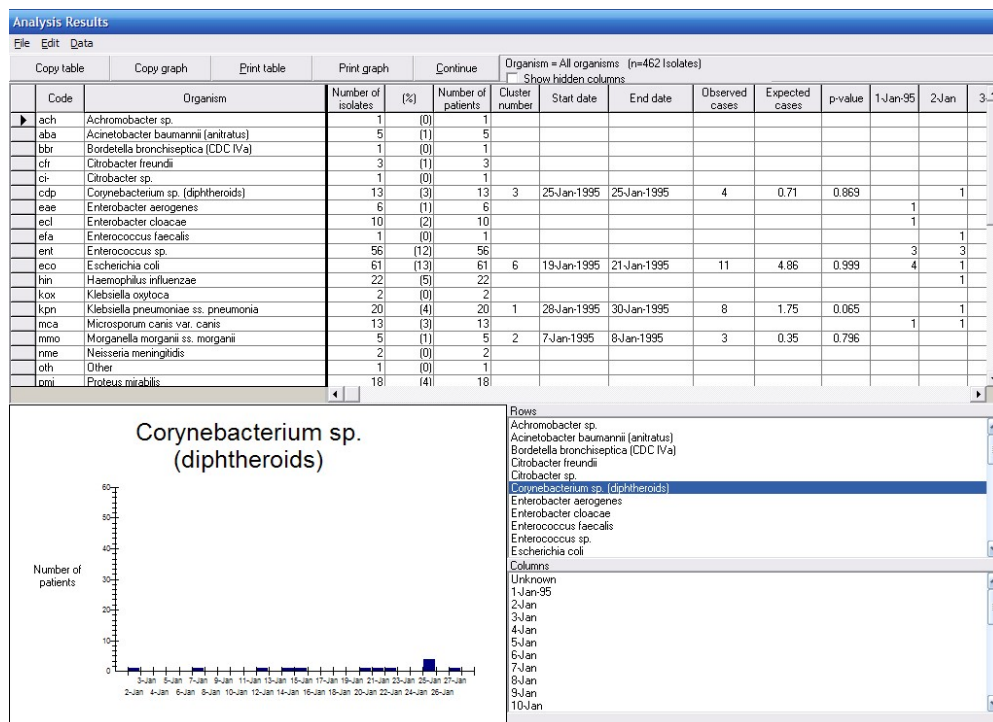
In addition to the typical columns for Isolate listing summaries, you see several additional columns in this output related to the clusters detected by SaTScan.

In the sample dataset with one month of data, WHONET found 39 bacterial species. SaTScan identified possible clusters in 7 of these species. Cluster number 1 is the cluster of greatest statistical significance (smallest p-value), *etc.* Only one of the identified clusters has a small p-value:

Organism	*Klebsiella pneumoniae*
Cluster dates	January 28-30
Observed	8 patients
Expected	1.75 patients
p-value	0.065

This particular organism is depicted in the below graph, confirming that there does indeed seem to be a rise in the number of patients with *K. pneumoniae* at the end of the month.

Klebsiella pneumoniae ss. pneumonia
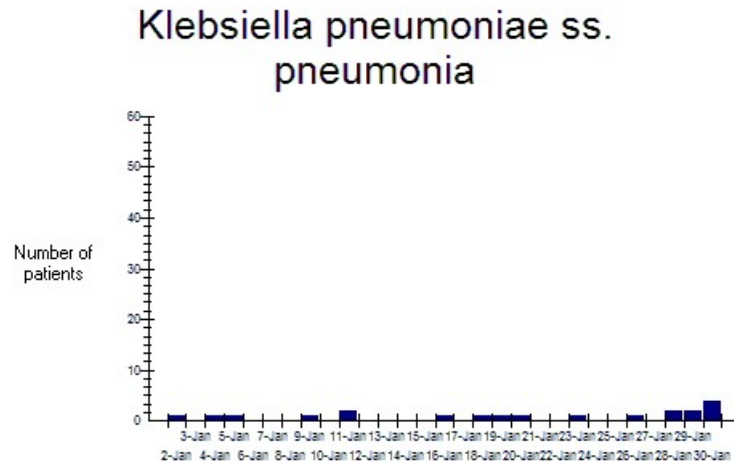
Part 2.   SaTScan and Resistance profile summaries

In Part 1, we saw how SaTScan statistics can be applied to organism frequency data. In this section, we will see how the same statistics can be applied to resistance profile data. For the output columns, put "Specimen date" by "Day". (As above, we are selecting "Day" in this tutorial because there is only one month of data.)

Go to "Analysis Type", and select "Resistance profiles". Check on the box for "Include SaTScan alerts".

For "One per patient", put "By patient", "First isolate only". For "Organisms", select *S. aureus*. Then begin the analysis. WHONET will first show you the Resistance profile listing. This is not affected by the SaTScan analysis. Hit Continue to proceed to the Resistance profile summary.

Analysis Results

File Edit Data

Copy table | Copy graph | Print table | Print graph | Continue | Organism = Staphylococcus aureus ss. aureus (n=60 Isolates) ☐ Show hidden columns

Letter = Resistant / Or intermediate / Space = Susceptible / · = Not tested

P = PEN S >= 29
E = ERY 14 - 22
C = CLI 15 - 20
O = OXA 11 - 12
G = GEN 13 - 14

V = VAN S >= 15
T = SXT 11 - 15
A = AMC S >= 20
R = CIP 16 - 20

Org	Profile	Resistance profile	Cluster number	Start date	End date	Observed cases	Expected cases	p-value	1-Jan-95	2-Jan	3-Jan	4-J
sau									1	1		
	──────────	·· ·· ·· ·· ·· ·· ·· ···										
	E	ERY										
	P	PEN									1	
	E R	ERY CIP										
	P R	PEN CIP	1	29-Jan-1995	30-Jan-1995	2	0.21	0.53				
	PE	PEN ERY	4	12-Jan-1995	12-Jan-1995	2	0.31	0.86			1	
	P─────V───	PEN ··· ··· ··· VAN ··· ··· ···										
	PE R	PEN ERY CIP								1		
	PE – R	PEN ERY ··· CIP									1	
	PEC R	PEN ERY CLI CIP										
	PECO A	PEN ERY CLI OXA AMC	3	17-Jan-1995	19-Jan-1995	2	0.28	0.8				
	PEC G T R	PEN ERY CLI GEN SXT CIP										
	PECO AR	PEN ERY CLI OXA AMC CIP									1	
	PE OG TAR	PEN ERY OXA GEN SXT AMC CIP										
	PECOG TAR	PEN ERY CLI OXA GEN SXT AMC CIP	2	13-Jan-1995	13-Jan-1995	2	0.28	0.8				

Profile: P   R

Number of patients — 60, 50, 40, 30, 20, 10, 0

26-Jan   27-Jan   28-Jan   29-Jan   30-Jan

Resistance profile

──────────
E
P
E   R
P   R
PE
P----V---
PE   R

Columns

Unknown
1-Jan-95
2-Jan
3-Jan
5-Jan
6-Jan
8-Jan
9-Jan
10-Jan

WHONET found 60 isolates of *S. aureus*, which cluster into 16 different resistance phenotypes. (The results "-" and "---" indicated that the indicated antimicrobials were not tested). From the 16 resistance phenotypes, SaTScan identified 4 possible clusters. Cluster number 1 has the following characteristics.

Resistance profile	PEN CIP	(non-susceptible only to PEN and CIP)
Cluster dates	January 29-30	
Observed	2 patients	
Expected	0.21 patients	
p-value	0.53	

The graph for this resistance profile is given above. Even though the number of isolates is very small, and the p-value is not, this cluster may still be of interest. In the entire month of January, this resistance profile was only seen three times, all at the end of month – once on January 26 (not detected by SaTScan), January 29, and January 30. In a further investigation, it would be useful to look at the patient locations. If you refer back to the Resistance profile listing, you will notice that the three isolates came respectively from location "xx = Unknown", "er = Emergency Room", and "op = Outpatient". So there is no epidemiologic link between these three patients with the available data, but one possibly would be revealed with further investigation of the patient medical records.

Part 3.   SaTScan – Cluster detection (Data Analysis)

In parts 1 and 2, we saw how SaTScan results can be integrated with standard WHONET analyses. In the next two parts, we will focus only on the information generated by SaTScan.

In "Analysis type", select "SaTScan – Cluster detection".   At the bottom of the screen, click on "SaTScan" to see additional options. In all above analyses, we used a Retrospective analysis. For this example, choose Prospective analysis.

For organisms, put "All organisms". For One per patient, put "By patient", "First isolate only". Begin the analysis.

Analysis Results

File Edit Data

Copy table	Copy graph	Print table	Print graph	Continue	Organism = All organisms (n=462 Isolates)

☐ Show hidden columns

|	Cluster number	Locations	Start date	End date	Observed cases	Expected cases	p-value
	1	kpn	28-Jan-1995	31-Jan-1995	8	2.36	0.057
	2	hin	28-Jan-1995	31-Jan-1995	6	2.6	0.858
	3	aba	25-Jan-1995	31-Jan-1995	3	1.23	0.999
	4	bsb	25-Jan-1995	31-Jan-1995	3	1.23	0.999
	5	cdp	20-Jan-1995	31-Jan-1995	8	4.86	0.999
▶	6	spn	26-Jan-1995	31-Jan-1995	2	0.76	0.999
	7	ent	29-Jan-1995	31-Jan-1995	8	5.27	1
	8	svi	17-Jan-1995	31-Jan-1995	8	5.37	1
	9	pr-	22-Jan-1995	31-Jan-1995	2	1	1
	10	pmi	28-Jan-1995	31-Jan-1995	3	1.89	1
	11	sau	25-Jan-1995	31-Jan-1995	17	14.21	1
	12	scn	24-Jan-1995	31-Jan-1995	25	22.16	1
	13	eco	17-Jan-1995	31-Jan-1995	32	29.28	1
	14	mca	28-Jan-1995	31-Jan-1995	2	1.54	1
	15	ecl	25-Jan-1995	31-Jan-1995	3	2.45	1
	16	dxe	19-Jan-1995	31-Jan-1995	2	1.66	1
	17	mmo	22-Jan-1995	31-Jan-1995	2	1.67	1
	18	pma	19-Jan-1995	31-Jan-1995	3	2.91	1

In Part 1, we saw that January had 39 bacterial species. Of the 39 species, SaTScan has identified 18 possible "ongoing" clusters, in other words clusters that potentially could include January 31, the final date of the analysis period. Of the 18 clusters, only one has a small p-value

Organism	*Klebsiella pneumoniae*
Cluster dates	January 28-31
Observed	8 patients
Expected	2.36 patients
p-value	0.057

This is the same cluster detected in Part 1 of this tutorial.

Part 4.   SaTScan – Space-and-Time cluster detection (Quick Analysis)

Now exit from the usual Data analysis program, and select "Quick analysis".
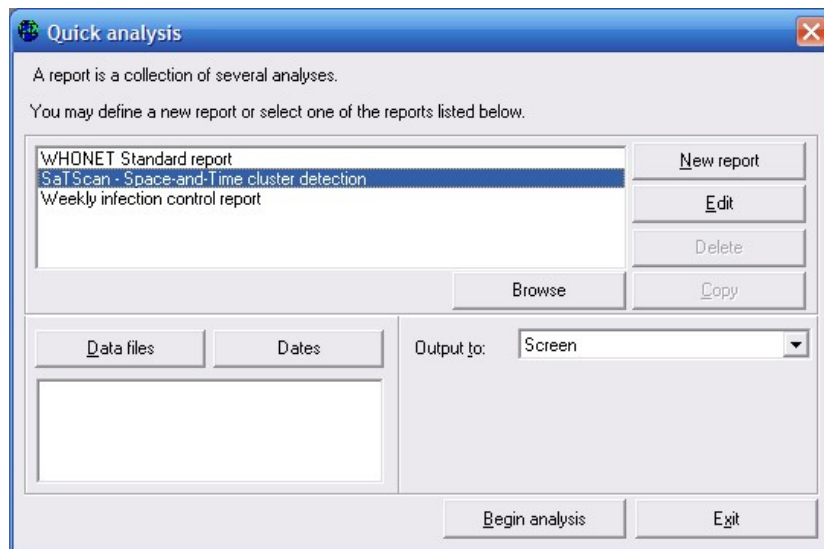
Choose the option "SaTScan – Space-and-Time cluster detection". If you click on "Edit", you will see the same SaTScan options menu which we saw in Data analysis.

For Data files, choose the w0195who.tst data file. Click on Begin analysis.

WHONET will display for you a two-part output. Part A – SaTScan report is the actual report prepared by the SaTScan software summarizing the details of the analysis performed and all statistical findings. This output begins with a brief summary of the analysis performed and the data summarized.

It then continues with a list of all of the identified clusters in order of decreasing statistical significance. Finally at the bottom of the output, you will see the detailed parameter settings used for the analysis.
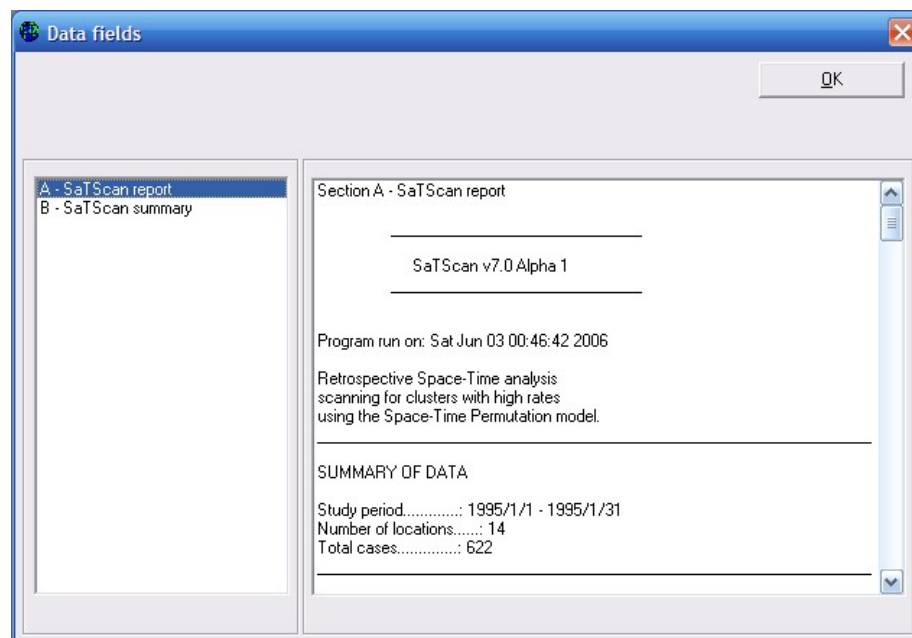
If you continue with Part B – SaTScan summary, you will see that WHONET has extracted the most important pieces of information for display to the user in a concise format. (Note: in this first version of WHONET 5.4, the formatting of the Quick Analysis results has not been optimized. This will be improved in the near future.)

Data fields

OK

A - SaTScan report
B - SaTScan summary

Section B - SaTScan summary

Cluster	Locations	Time frame	Observed	Expected	p-value
1	su	3-Jan-1995-4-Jan-1995	11	1.85	0.002
2	op	25-Jan-1995-26-Jan-1995	29	11.76	0.002
3	wn	23-Jan-1995-23-Jan-1995	7	0.95	0.004
4	er	22-Jan-1995-22-Jan-1995	12	3.01	0.005
5	4w	7-Jan-1995-7-Jan-1995	9	1.85	0.007
6	hn	20-Jan-1995-21-Jan-1995	9	2.09	0.023
7	iw	27-Jan-1995-27-Jan-1995	3	0.14	0.023
8	ie	14-Jan-1995-14-Jan-1995	6	1.17	0.16
9	xx	9-Jan-1995-10-Jan-1995	3	0.25	0.214
10	hs	17-Jan-1995-17-Jan-1995	11	3.72	0.228
11	4b	1-Jan-1995-5-Jan-1995	19	8.92	0.321
12	4e	13-Jan-1995-18-Jan-1995	13	5.39	0.55
13	ws	12-Jan-1995-12-Jan-1995	6	1.84	0.938

Kulldorf M. and Information Management Services, Inc. SaTScan(TM) v.7.0: Software
SaTScan(TM) is a trademark of Martin Kulldorf. The SaTScan(TM) software was deve