

A framework for Computer Supported Outbreak Detection

Maria Grünewald

Baki Cakici, Kenneth Hebing, Anette Hulth and Paul Saretok



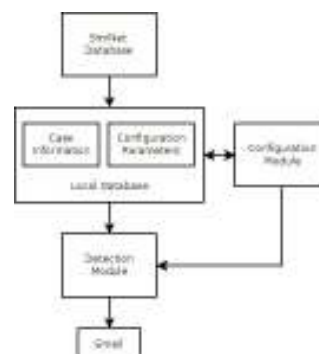
What is CASE

- System to automatically detect disease outbreaks
- Developed and used at SMI
- Runs every night analyzing data from the SMI database
- Signals based on statistical analysis
- Source code is licensed under GNU General Public License



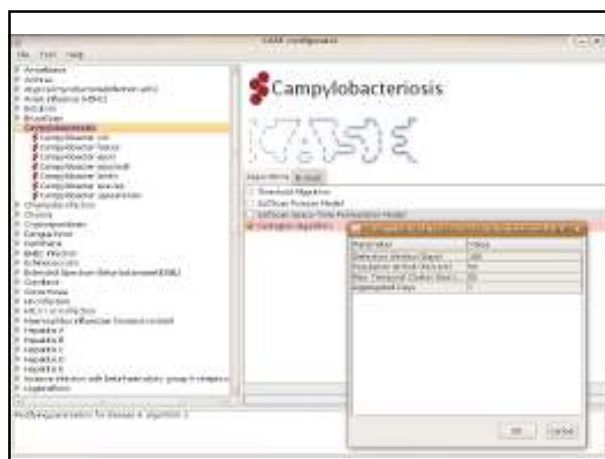
The SMI database

- National notifiable disease database
- Physicians report infectious diseases that are notifiable by law
- Most reports are electronic
- Incoming reports are checked and registered



User interface

- Graphical user interface
- Only the system administrator uses the interface
- The end user will only see the emails sent by the system



What are we looking for?

- What is an outbreak?
- Deviations from normal rate of disease

Specifying to the computer what a deviation is:

- Threshold?
- Increase over time?
- Spatial deviations?
- In relation to historic data?



HEALTH PROTECTION AGENCY

Threshold algorithm

- Gives signal if number of cases exceeds threshold during a detection window

Example: More than 4 cases in two weeks

- Threshold and detection window can be specified for each disease (or subtype) individually



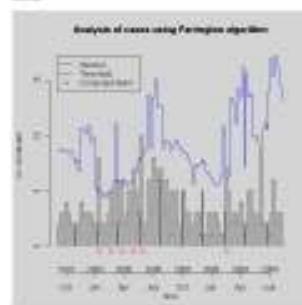
HEALTH PROTECTION AGENCY

Farrington algorithm

- Compares number of cases with historical data
- Model takes season into account
- Used, and developed, at the Health protection Agency in England and Wales
- CASE parser gives a signal only if deviation occurred during last 7 days, or during the seven days before that



HEALTH PROTECTION AGENCY



The graph shows the observed cases in this period in relation to the expected range.

Why does the Farrington algorithm (Farrington et al. 2005) work?

The algorithm compares the observed number of cases in a given week against the expected number of cases based on historical data. The expected number is calculated by the model. Deviations from the expected number are identified by the model. Deviations from the expected number are identified by the model. Deviations from the expected number are identified by the model.

Farrington D, Kulkarni T, Beale R, et al. (2005) A statistical algorithm for the detection of outbreaks of infectious disease. *Journal of the Royal Society Interface*, *2*, 125-136, 2005.

The full reference can be found in the references.

SaTScan Space-Time Permutation

(Kulldorff et al. 2005)

- Detects deviations that occur in time and space simultaneously
- Not restricted to regional boundaries

Example signal: Incidence increases at Gotland

Examples no signal:

Gotland always has high incidence

Seasonal variation that affects all Sweden



HEALTH PROTECTION AGENCY

SaTScan Poisson

- Compares different locations with each other during specified time window
- Is not restricted to regional boundaries
- Takes population density into account

Example: During the last month cases has been more frequent in an area near Malmö than the rest of the country



HEALTH PROTECTION AGENCY

Add algorithms yourself

- Add your favorite algorithm!

The system is not restricted to the algorithms listed here but is a framework where algorithms can be added



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

Can CASE replace the epidemiologist?

- No!
- The intuition, knowledge and flexibility of a human expert can not be replaced
- Data may be available from multiple sources to the epidemiologist
- The intention is to assist, not replace



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

So how can CASE help me?

- May detect patterns that a human expert would overlook
- Potentially earlier detection
- Aid less experienced epidemiologist
- ...



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

Can I use CASE?

- GNU General Public License Version 3
- Available for download at <http://smisvn.smi.se/case>
- Runs on Linux and Windows
- May require adaptation to database



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

Challenges

- Getting good quality data fast
- Settings for parameters
- Feed-back data



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

Future work

- More feedback from epidemiologists
- Build in possibility to run many dates at once (for evaluation purposes)
- Comparison with actual outbreaks (sensitivity, specificity, timeliness)
- Evaluation with simulated data



SMI
STRENGTHENING
MISVNS
INTEGRATED
EVIDENCE
FOR
PUBLIC
HEALTH

Summary

- CASE is a framework for computer supported outbreak detection
- CASE is running, and we are working on extensions
- CASE is available as open source and we are happy if you use it!



References

- CASE-A Framework for Computer Supported Outbreak Detection, Baki Cakici, Kenneth Hebing, Maria Grünewald, Paul Saretok and Anette Hulth (submitted to journal)
- Farrington CP, Andrews NJ, Beale AD, Catchpole MA: A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 1996, 159(3):547-563.
- Kulldorff M: A spatial scan statistic. *Communications in Statistics: Theory and Methods* 1997, 26:1481-1496.
- Kulldorff M, Heffernan J Hartman, Assunção R, Mostashari F: A Space-Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med* 2005, 2(3):e59.
- SaTScan - Software for the spatial, temporal, and space-time scan statistics [<http://www.satscan.org>].

