T-Cube Web Interface as a tool for detecting disease outbreaks in in real-time: A pilot in India and Sri Lanka

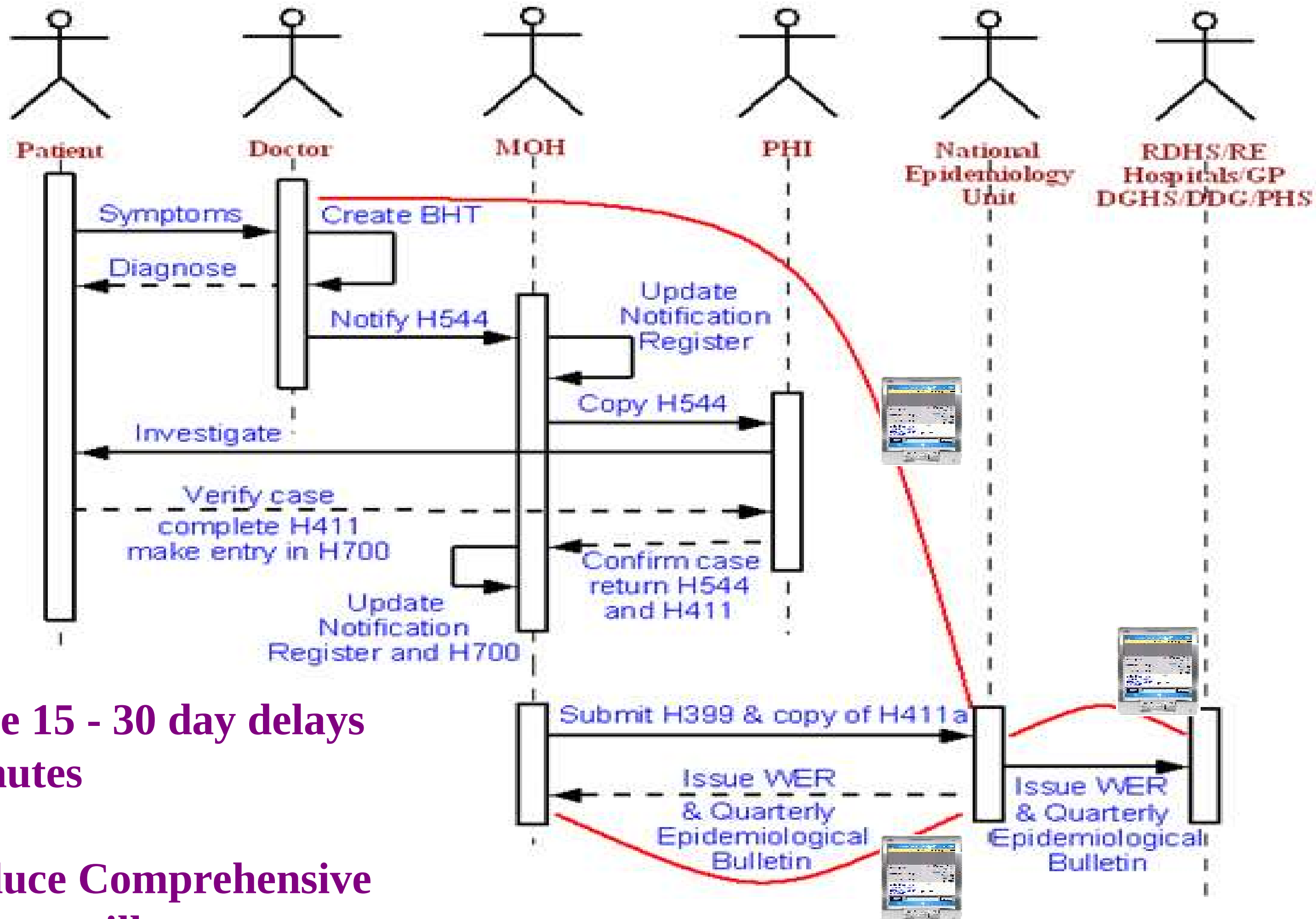
IEEE RIVF International Conference on Computing and Communication Technology
Research, Innovation, and Vision for the Future

November 02, 2010
Vietnam National University, Hanoi, Vietnam

Nuwan Waidyanatha
LIRNE*asia*
Email: nuwan@lirneasia.net
http://www.lirneasia.net/profiles/nuwan-waidyanatha
Mobile: +8613888446352 (cn) +94773710394 (lk)

This work was carried out with the aid of a grant from the International Development Research Centre, Canada.

Problem the RTBP is trying to solve in developing countries



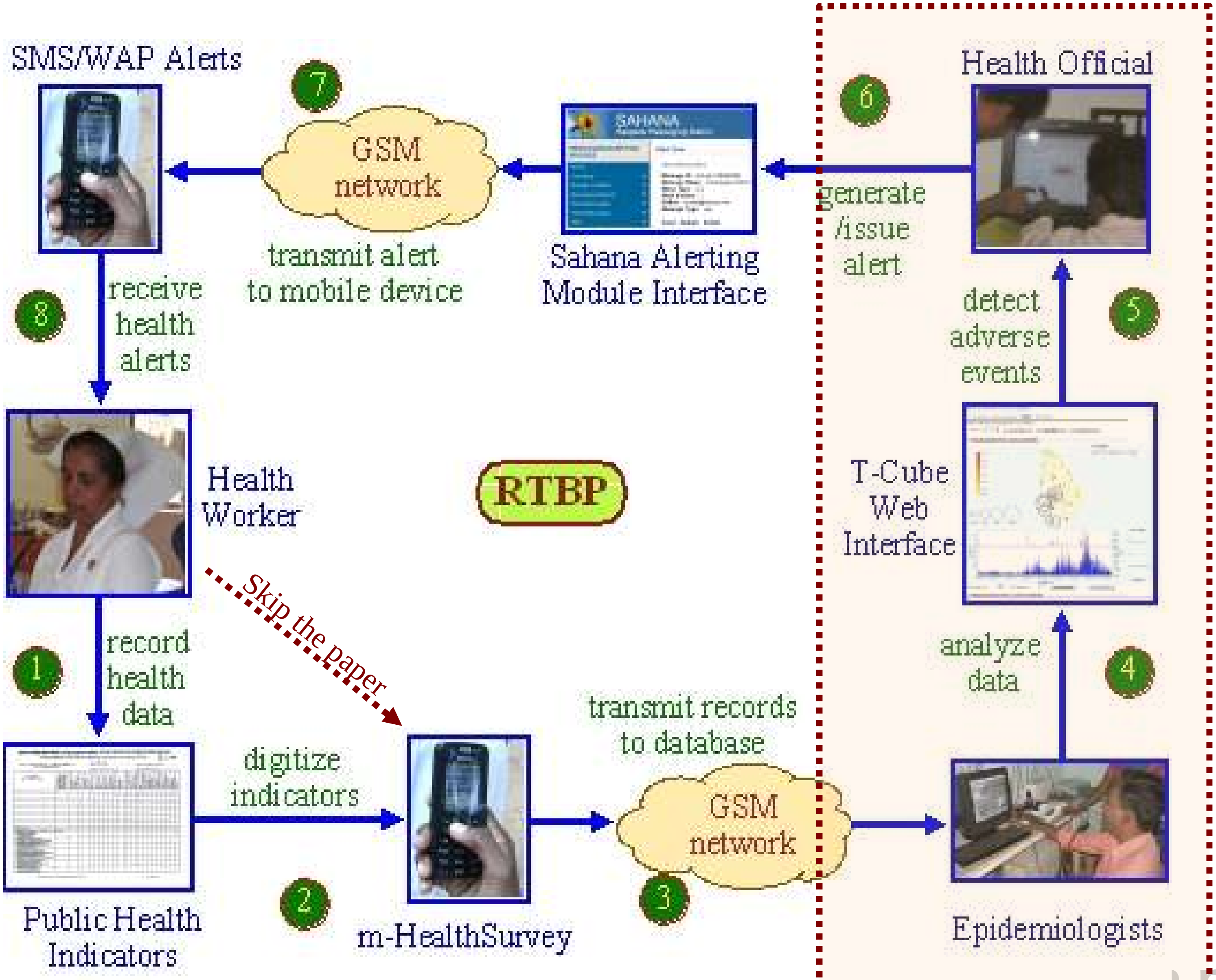
Reduce 15 - 30 day delays to Minutes

Introduce Comprehensive active surveillance

Black arrows: current manual paper/postal system for health data collection and reporting

Red lines: RTBP mobile phone communication system for heath data collection and reporting

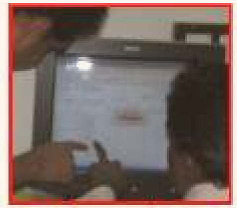
Data collection, Event detection, and Situational-wareness/Alerting



Research Question: “*Can software programs that analyze health statistics and mobile phone applications that send and receive the health information potentially be effective in the early detection and mitigation of disease outbreaks?*”

Health workers digitize health records with mHealthSurvey

GSM Network



Epidemiologist analyze data with T-Cube Web Interface

GSM Network

Epidemiologist issue health alerts on to Health worker phones using Sahana Alerting

Data Collection

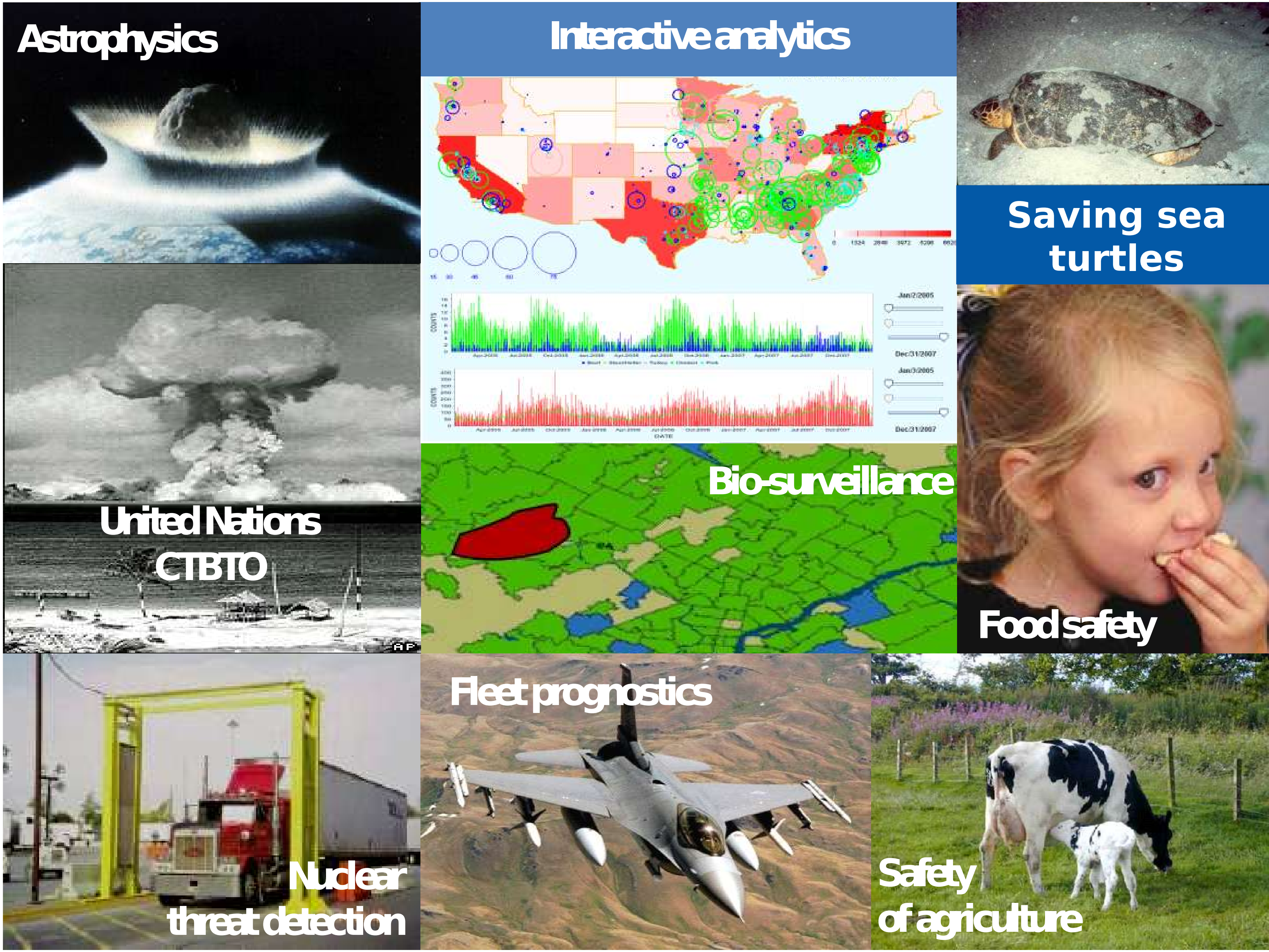
mHealthSurvey is a data entry software that works on any standard java–enabled mobile phone. A typical record contains the patient visitation date, location, gender, age, disease, symptoms, and signs. Data is transmitted over GPRS cellular networks.

Event Detection

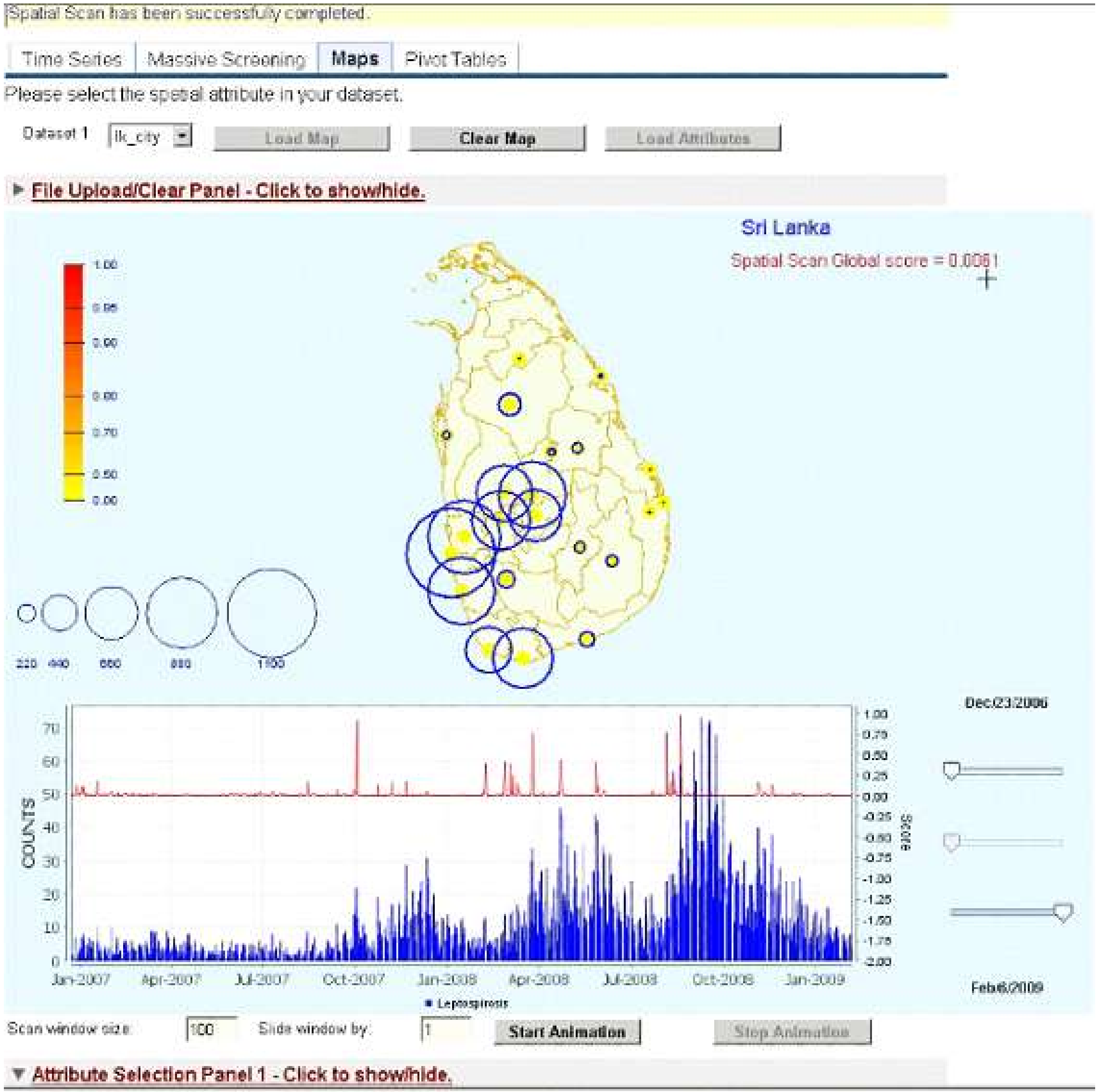
T-Cube Web Interface (TCWI) is an Internet browser based tool to visualize and manipulate large spatio-temporal data sets. Epidemiologists can pin down a potential outbreak of, for instance, a gastrointestinal disease among children in the Sevanipatti PHC health division.

Alerting

Sahana Alerting Module (SAM) allows for the generic dissemination of localized and standardized interoperable messages. Selected groups of recipients would receive the single-entry of the message via SMS, Email, and Web.


Astrophysics
Interactive analytics
Saving sea turtles
United Nations
CTBTO
Bio-surveillance
Food safety
Fleet prognostics
Nuclear
threat detection
Safety
of agriculture

T-Cube Web Interface (TCWI) by *Auton Lab*



- AD Tree data structure
- Trained Bayesian Networks
- Fast response to queries
- Statistical estimations techniques
- Data visualization over temporal and spatial dimensions
- Automated alerts

Pre-Screening using Massive Tempotal Scan

Minimum Support 0

Most significant window only

Last Day Only

Start Date 09/22/2009

End Date 08/31/2010

Period of Reference

Start Date 09/22/2009

End Date 08/31/2010

Scan Option: Upper Tail

Baseline: All Data

Reference Type: All Past

Reference Window: 28 Lag: 0

Use Drill-Down

Use Adaptive Baseline

Run Screening

Screening Results

The following screenings are available:

- Escalating Fever Diseases
- Escalating Non-communicable Diseases
- Escalating Notifiable Diseases
- Escalating other communicable Diseases

This screening shows data about diseases classified as high-impact and notifiable. It will show instances of these disease that have been statistically unusual recently.(21 queries)

Id	Query	Date	PValue	Window	Count	Expected Count	Ranked PValue	Alt Date	Alt PValue	Alt Windov
1	disease=(Chicken_Pox),age_grp=(15-20),dis_priority=(Notifiable)	08/31/2010	3.798E-5	21	4	0.06	2.899E-3			
2	disease=(Chicken_Pox),age_grp=(15-20),location=(Sandalankawa),dis_priority=(Notifiable)	08/31/2010	3.094E-4	21	3	0	2.899E-3			
3	age_grp=(15-20),location=(Sandalankawa),dis_priority=(Notifiable)	08/31/2010	2.788E-3	21	3	0.14	2.899E-3			
4	location=(Sandalankawa),dis_priority=(Notifiable)	08/31/2010	4.694E-3	21	4	0.51	2.899E-3			
5	disease=(Chicken_Pox),location=(Sandalankawa),dis_priority=(Notifiable)	08/31/2010	5.295E-3	21	3	0.22	2.899E-3			
6	age_grp=(15-20),dis_priority=(Notifiable)	08/31/2010	1.000E-2	28	9	3.26	2.899E-3			

On 08/31/2010, we note that the counts in last 21 days are significantly higher than expectation. The chance of seeing anything more extreme is about one in 26,000 cases.

Prev 10 | Next 10

Dump Query | Use For Drill-Down | Save To List | Name: Escalating Notifiable Disea | Save Results to File | Print List

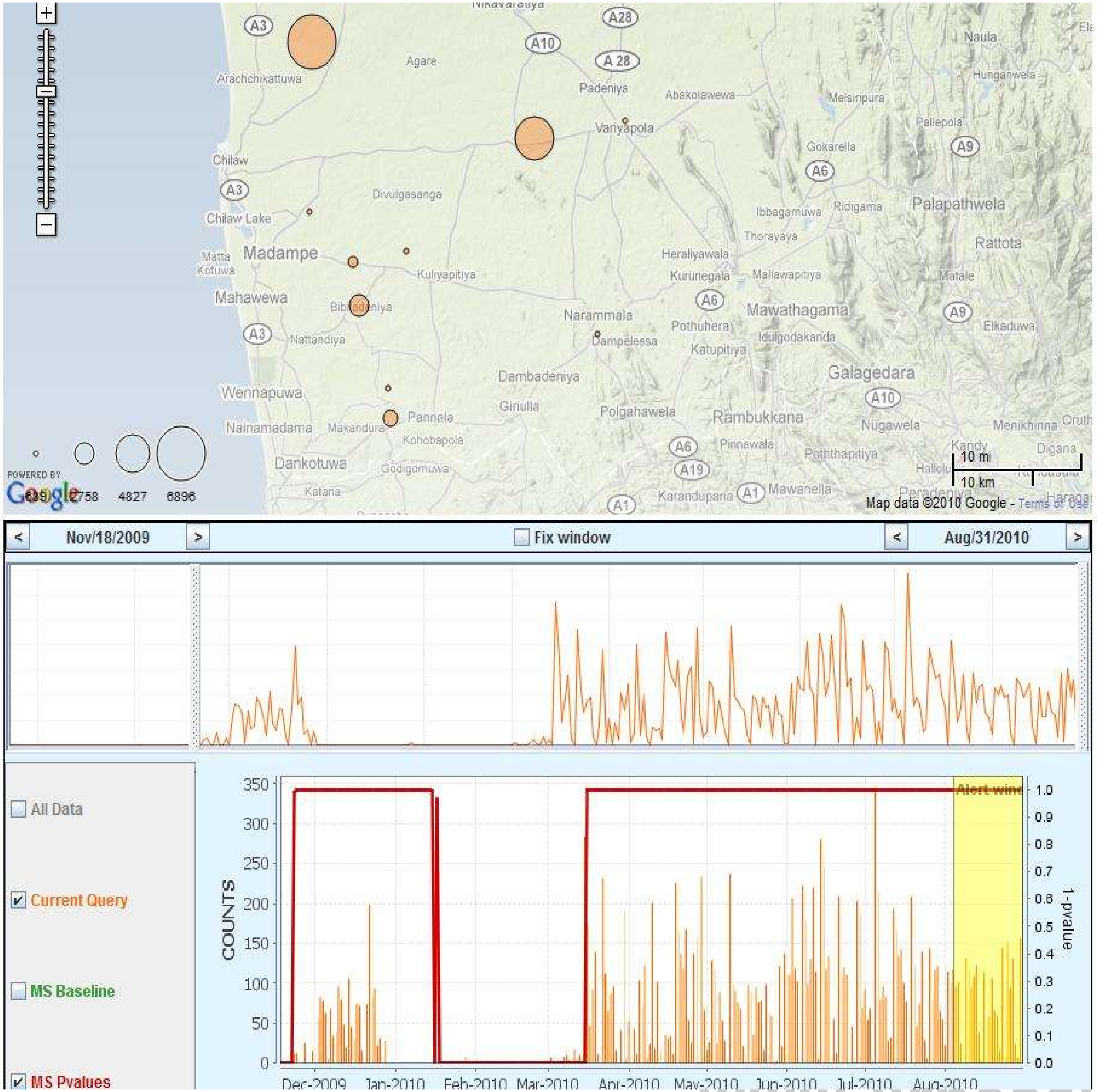
Rate usefulness of the result: No Answer 1 2 3 4 5

User Name

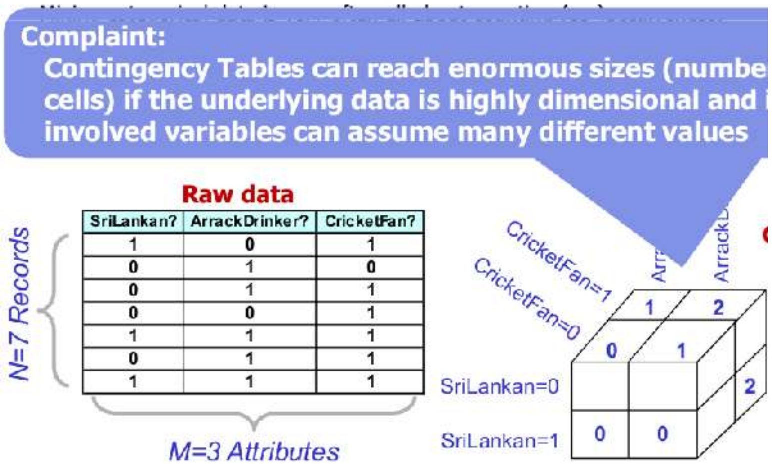
Comments

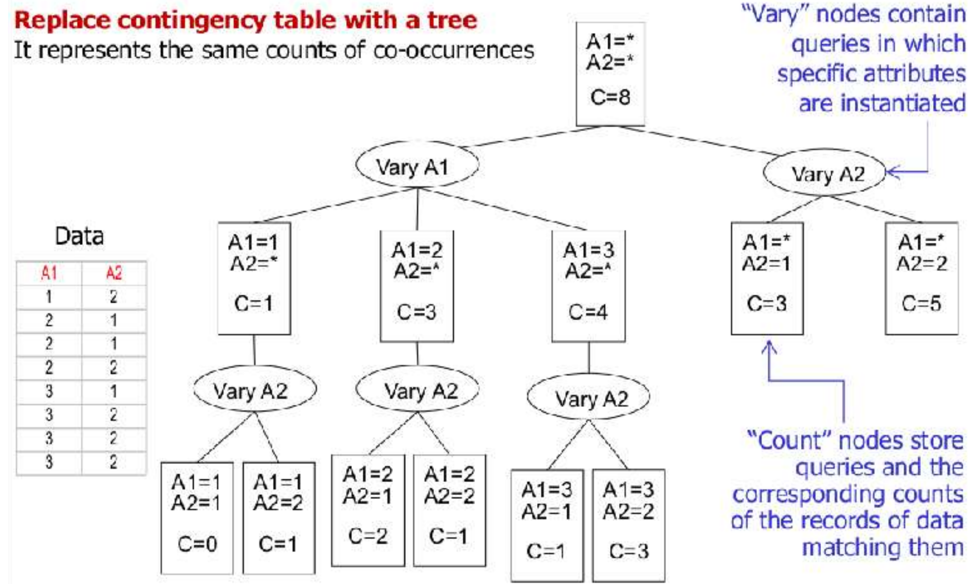
Submit Feedback

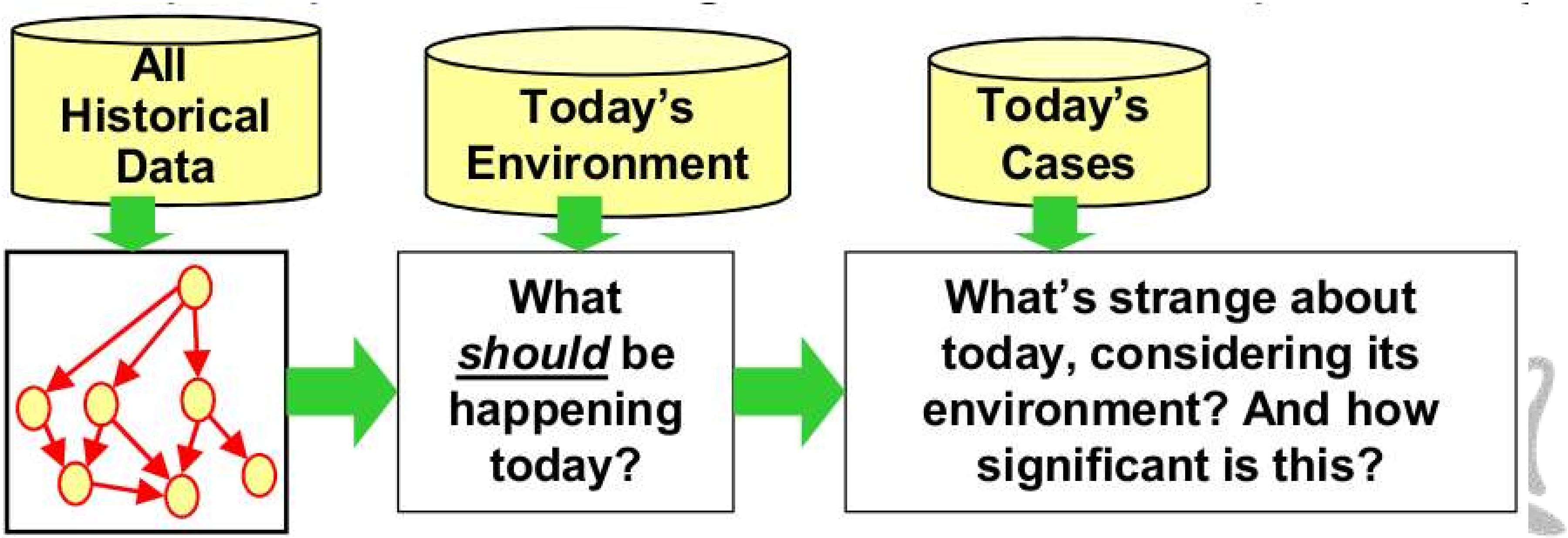
T-Cube Web Interface – Spatio – Temporal Presentation



Overview of the T-Cube data structure and computations







Research design

- 24 Health Sub Center Village Nurses
- 4 Public Health Center Sector Health Nurses, Health Inspectors, and Data Entry Operators
- 1 Integrated Disease Surveillance Program Unit of the Deputy Director of Health Services
- Thirupathur Block, Sivagangai District, Tamil Nadu, India

- 12 District/Base Hospitals and Clinics
- 15 Sarvodaya Suwadana Center Assistants
- 4 Medical Officer of Health divisions & 1 Regional Epidemiology Unit
- Kurunegala District, Wayamba Province, Sri Lanka

Evaluation of TCWI

- **Replication study** :: Sri Lankan Weekly Epidemiological Return (WER) reports published at www.epid.gov.lk notifiable disease counts tabulated by District was semi synthesized by distributing the weekly counts as daily counts taking day-of-week effect, gender distribution, and age representations.
- **Study the reliability and effectiveness** :: significant events detected by T-Cube is compared with the ground truth and also weighed on the response actions or inaction
- **Competency exercise** :: injected fake data over a period of 5 days and the subjects, unaware of the prefabricated events, were asked to detect most significant events
- **T-Cube Acceptance** :: a questionnaire was designed based on the Technology Assessment Methodology (TAM) and was subject to TCWI users as well as health official associated with T-Cube who make decisions on whether or not to take action
- **Cost analysis** :: compare the economic efficiencies and cost effectiveness between present detection/analyses system and T-Cube

Replication study using synthesized WER data

- We took 2007 – 2009 Weekly Epidemiological Returns publicly available data - http://www.epid.gov.lk/
- Synthesized the data to match that similar to the RTBP dimensions by distributing the district weekly aggregates

- day-of-week visitation densities (M - F)

- female to male ration

- age-groups (0-5, 6-14, 15-20, 21-45, 46-65, above 65)

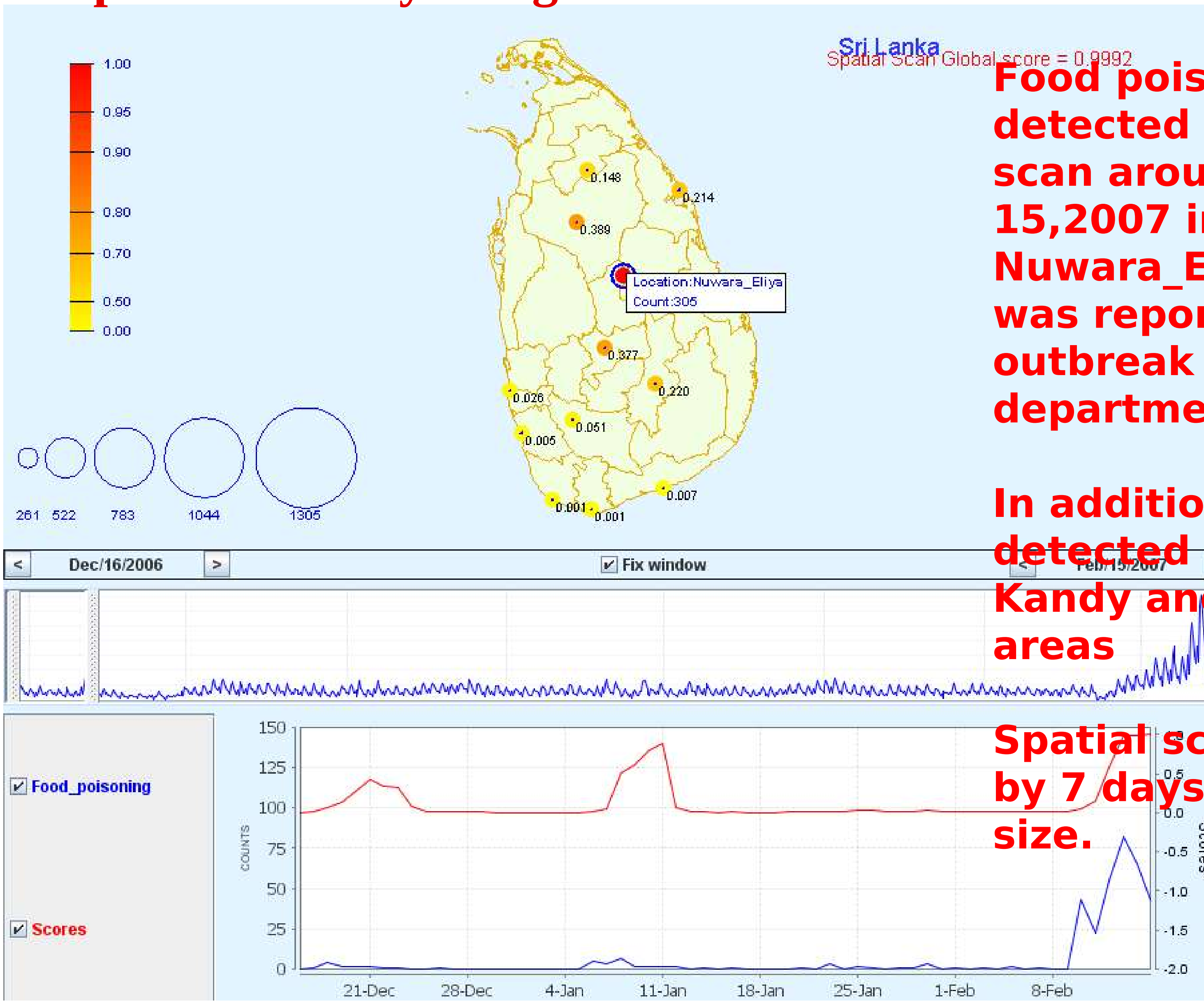
WER Sri Lanka – Vol. 37 No. 36 04th – 10th September 2010

Table 4: Selected notifiable diseases reported by Medical Officers of Health

28th August - 03rd September 2010(34th Week)

DPDHS Division	Dengue Fever / DHF*		Dysentery		Encephalitis		Enteric Fever		Food Poisoning		Leptospirosis		Typhus Fever		Viral Hepatitis		Human Rabies		Returns Re-
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	%
Colombo	111	5107	2	226	0	14	1	101	0	32	15	424	0	7	0	48	0	1	**92**
Gampaha	47	3502	3	119	0	19	0	36	1	19	13	295	0	12	2	76	0	4	**67**
Kalutara	26	1579	1	183	0	13	0	17	0	74	4	248	0	2	1	28	0	1	**58**
Kandy	37	1436	1	243	0	4	0	22	1	6	3	80	0	111	8	98	0	1	**74**
Matale	5	541	2	257	0	5	0	30	0	70	6	78	0	5	1	41	0	0	**67**
Nuwara	13	189	6	296	0	0	1	102	0	84	0	21	1	51	1	33	0	0	**92**
Galle	45	976	2	205	0	5	0	5	1	13	3	68	1	19	0	11	0	3	**95**
Hambanto	13	703	2	63	0	6	0	1	0	10	0	76	1	71	0	9	0	0	**73**
Matara	22	515	2	146	0	8	0	9	0	49	19	244	4	108	0	17	0	0	**88**

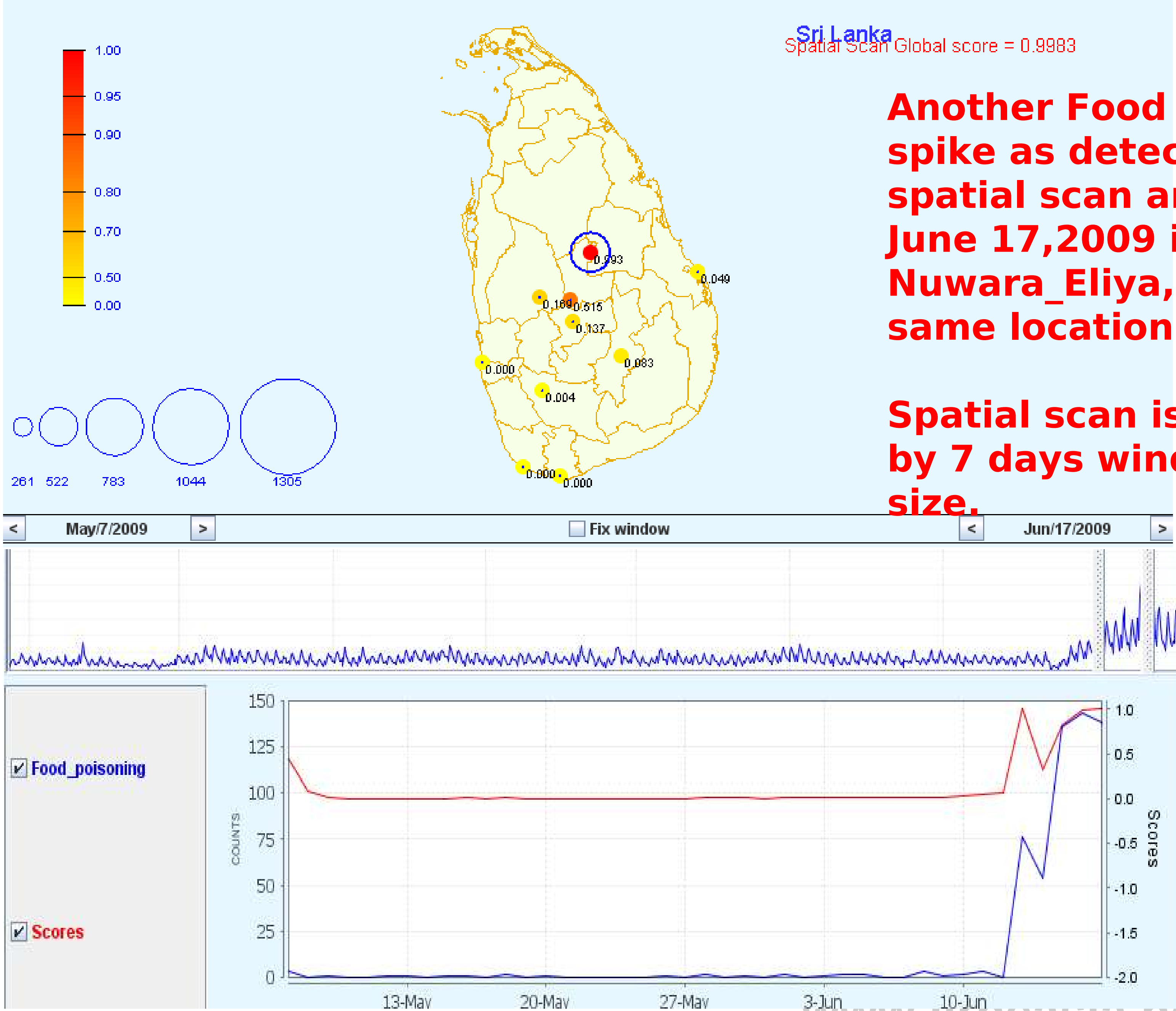
Replication study using Sri Lanka WER data 2007 - 2009



Food poison spike as detected by spatial scan around Feb 15,2007 in Nuwara_Eliya, which was reported as outbreak by health department.

In addition TCWI detected spikes in Kandy and Vauvniya areas

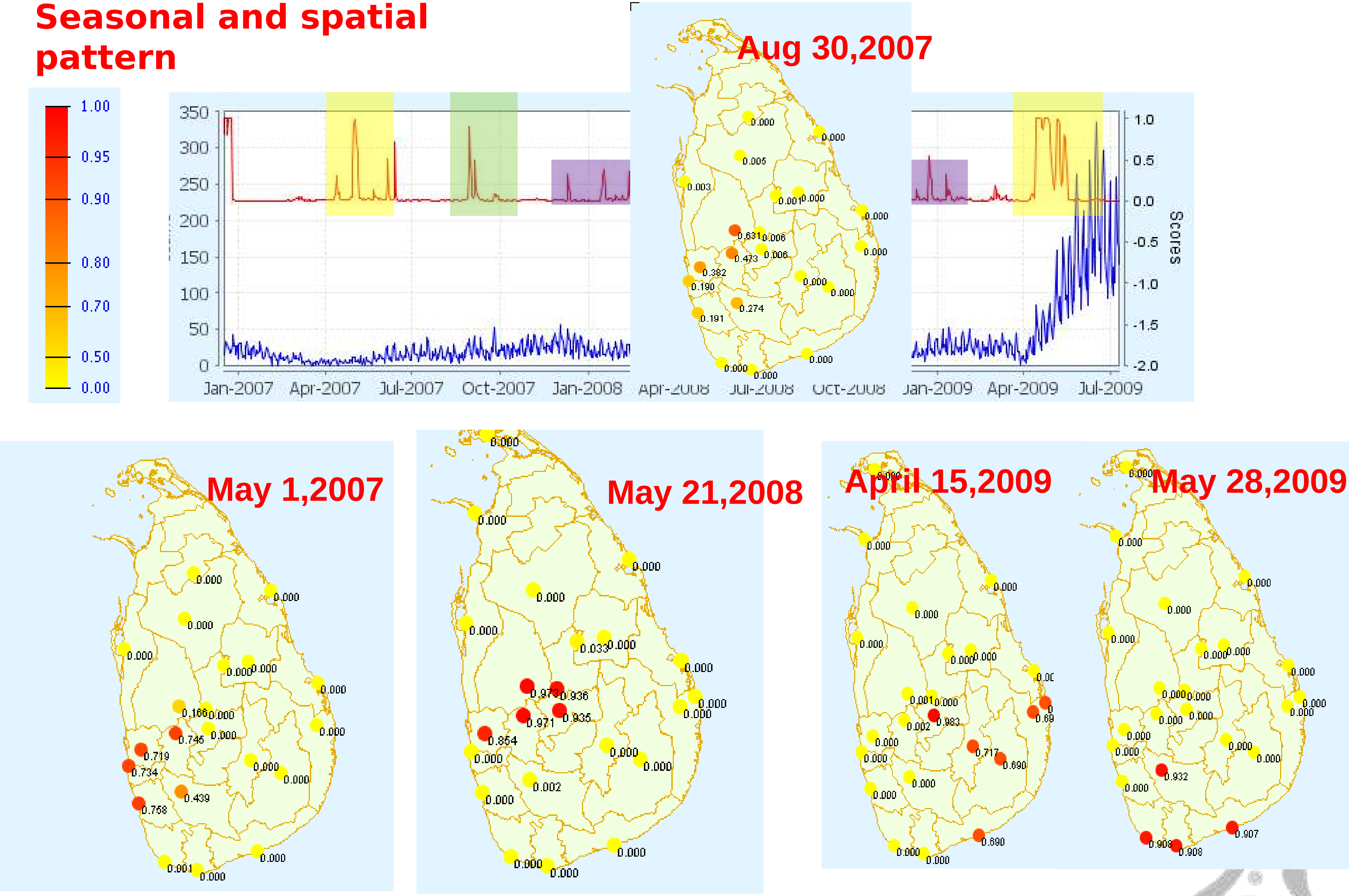
Spatial scan is run by 7 days windows size.



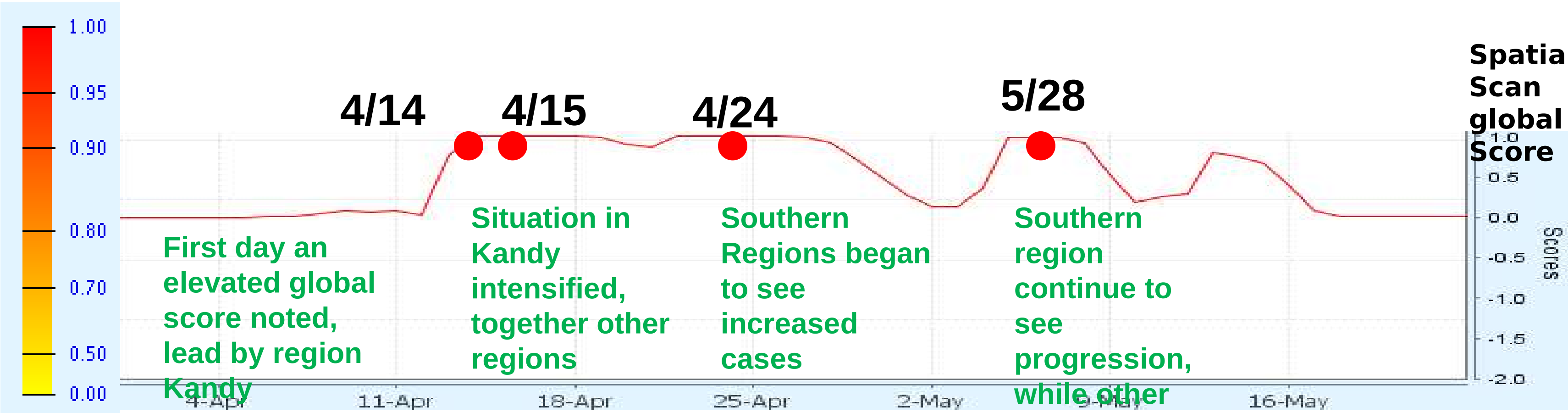
Another Food poison spike as detected by spatial scan around June 17,2009 in Nuwara_Eliya, the same location.

Spatial scan is run by 7 days windows size.

Dengue Fever Seasonal and spatial pattern



Progression of Dengue Fever outbreak in April – June 2009

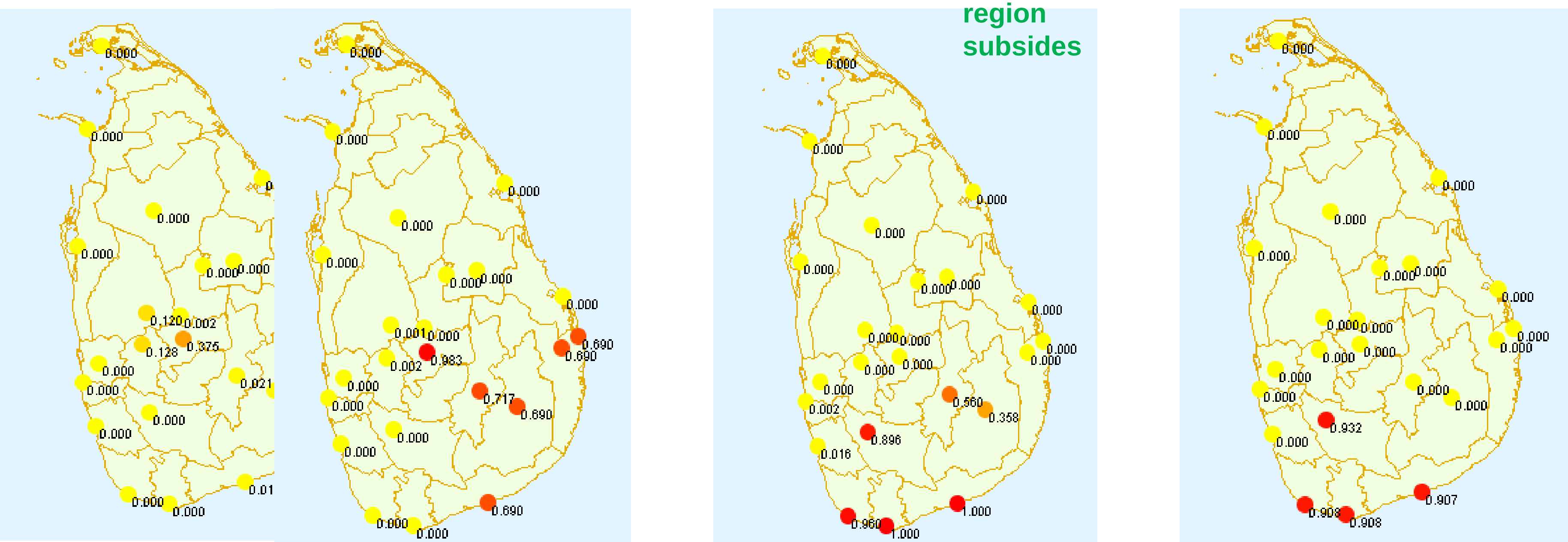


First day an elevated global score noted, lead by region Kandy

Situation in Kandy intensified, together other regions

Southern Regions began to see increased cases

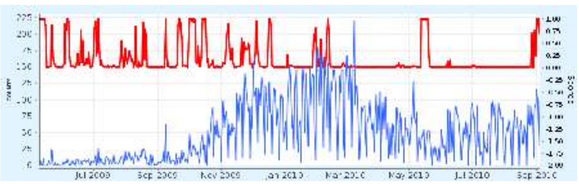
Southern region continue to see progression, while other region subsides



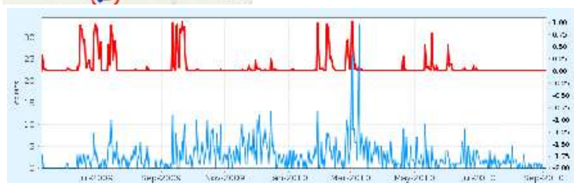
Most frequently occurring wide spreading infectious disease outbreaks

These findings are from TCWI's spatial scan algorithms

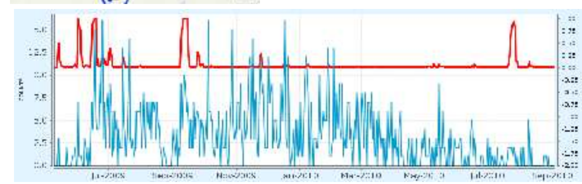
Common Cold, Sivaganga District – India, 18 outbreak episodes to date with over 23,188 cases.



Worm Infestation, Sivaganga District – India, 13 outbreak episodes to date with over 1,236 cases.

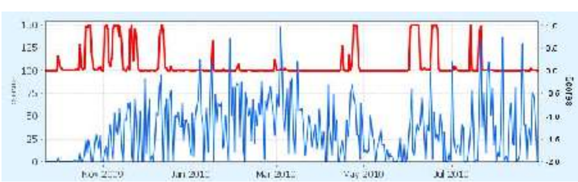


Dysentery, Sivaganga District – India, 5 outbreak episodes to date with over 1,541 cases.

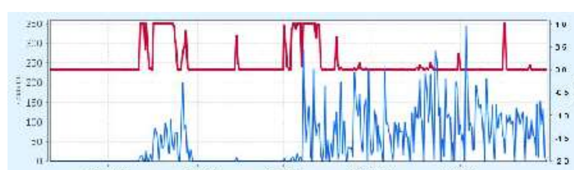


Common cold is the most popular but gastrointestinal infectious are, relatively, the most visible

Cough, Kurnegala District – Sri Lanka, 11 outbreak episodes to date with over X cases.



Respiratory Tract Infection, Kurnegala District – Sri Lanka, 09 outbreak episodes to date with over X cases.

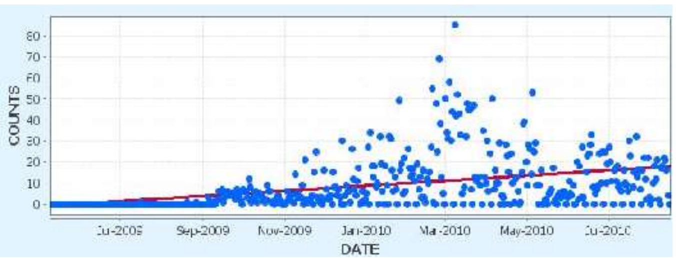


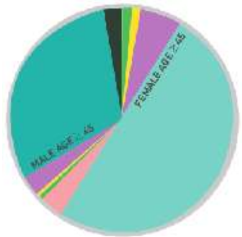
Tonsilitis, Kurnegala District – Sri Lanka, 07 outbreak episodes to date with over X cases to date

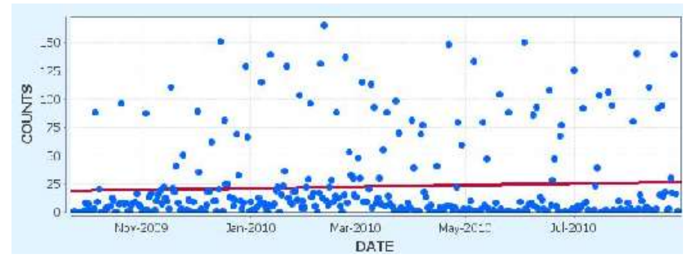
Respiratory infectious diseases, a correlated with environmental factors, are the most common

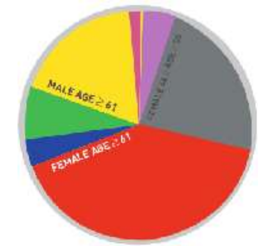
Trends in selected noncommunicable disease

These findings are from TCWI's statistical estimation and pivot table analysis methods

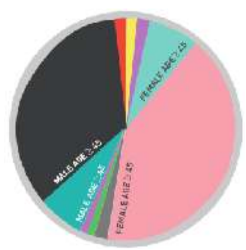


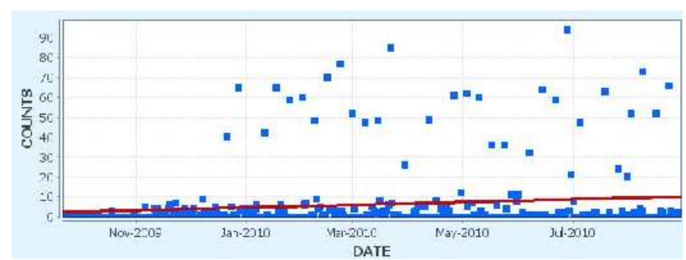


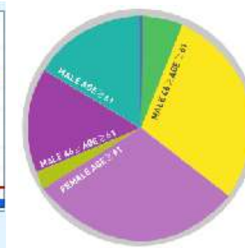




Hypertension (High Blood Pressure) has a linearly increasing trend over the one year period in both countries with Females and Males over 45 years of age showing to be the most vulnerable. The dtrend in India shows an unusual increase between March and May 2010; while the reported cases are consistent throughout the year in Sri Lanka.





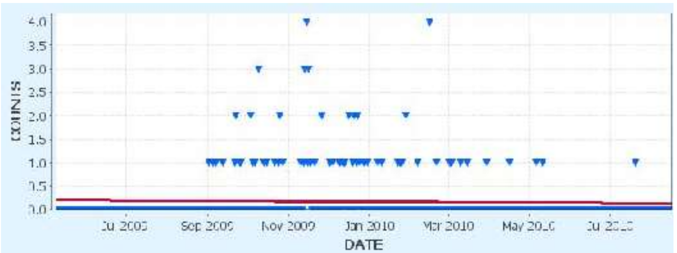


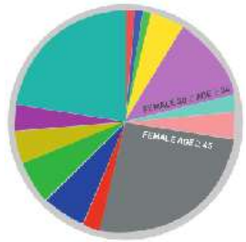
Diabetes-Mellitus has a linearly increasing trend over the one year period in both countries with Indians over 40 years of age and Sri Lankan over 45 years of age to be the most vulnerable groups.

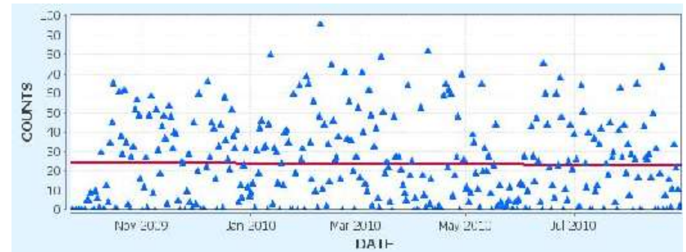
Given that the Male to Female ratios, approximately, in Tamil Nadu, India and Kurunegala, Sri Lanka are both 1 : 1; statistics to date show females to be more susceptible to the above mentioned life style diseases.

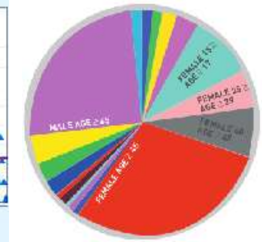
Trends in selected noncommunicable disease

These findings are from TCWI's statistical estimation and pivot table analysis methods

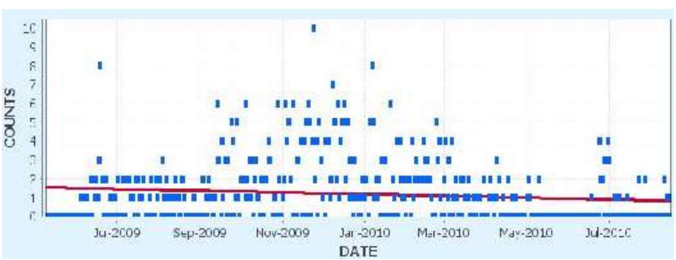


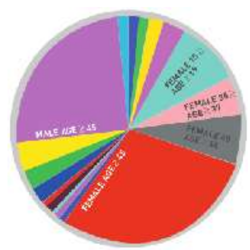


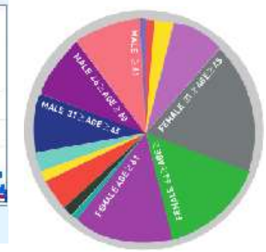




Arthritis and Rheumatoid-Arthritis has a linearly stagnate trend over the one year period in both countries with Males over 45 years of age and Females over 35 years of age to be the most susceptible in India; similarly Males over 45 and Females over 31 years of age to be the most vulnerable groups.



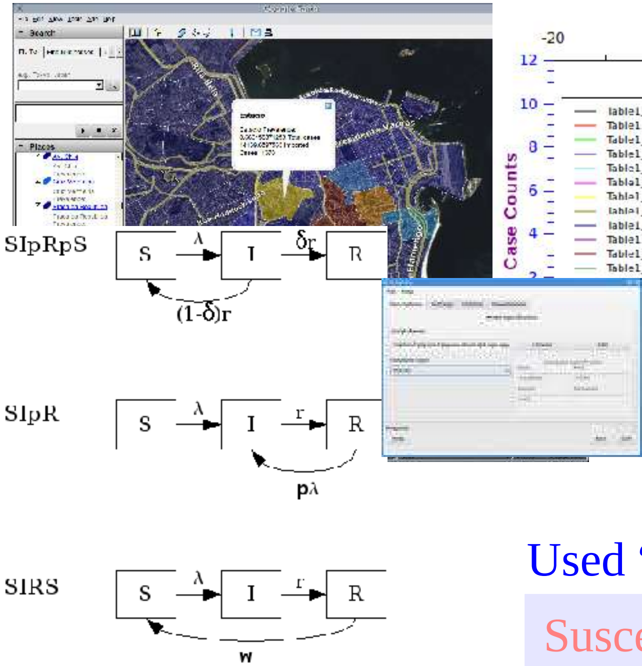


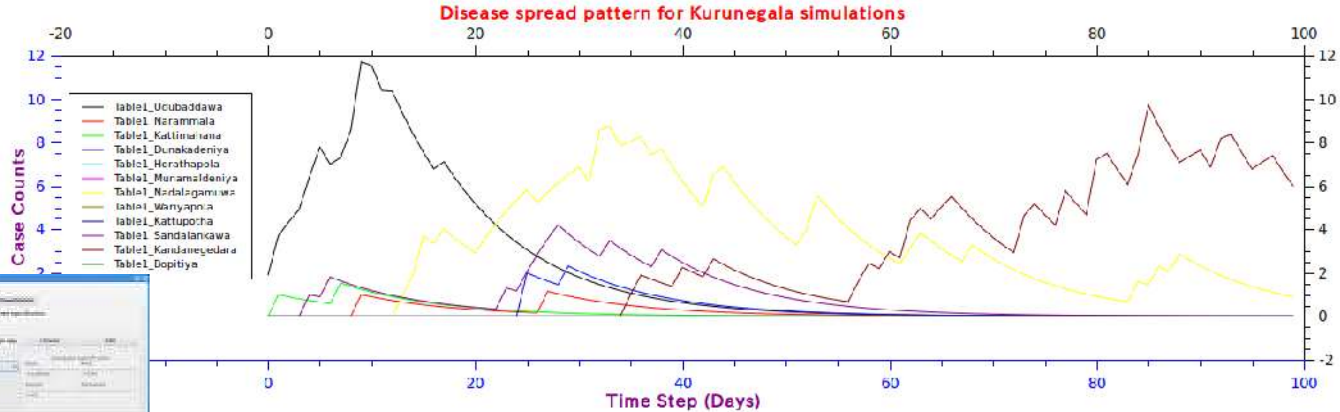


Asthma has a linearly decreasing trend over the one year period in both countries; the dtrend shows the counts to increase during the rainy season, India: Sept'09-Jan'10 and Sri Lanka: Nov '09-Jan '10. In India, only males over 45 years of age are affected but females in all age groups are affected. Both Male and Female over 31 years of age are in Sri Lanka are equally vulnerable.

Given that the Male to Female ratios, approximately, in Tamil Nadu, India and Kurunegala, Sri Lanka are both 1 : 1; statistics to date show females to be more susceptible to the above mentioned life style diseases.

TCWI Competency Assessments with Injected Synthetic data

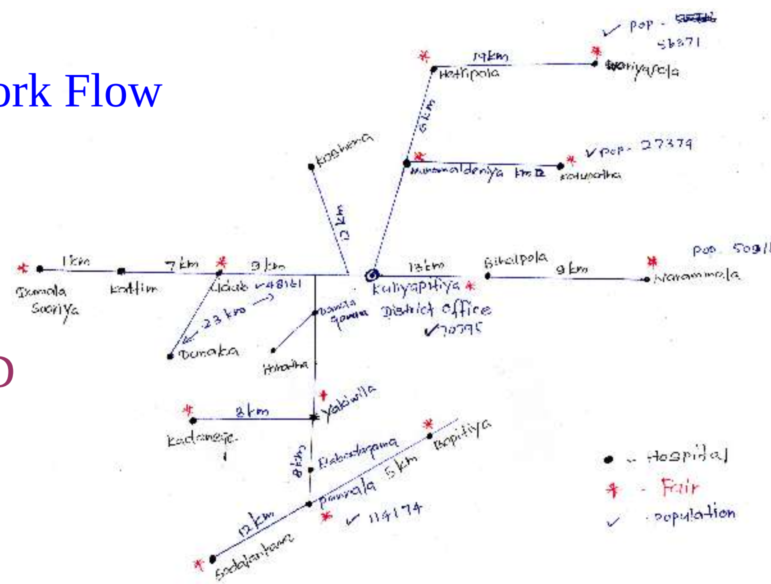




Used “**Epigrass**” to generate synthetic data with a SEIR model

Susceptible → Exposed → Infected → Recovered

With a Network Flow

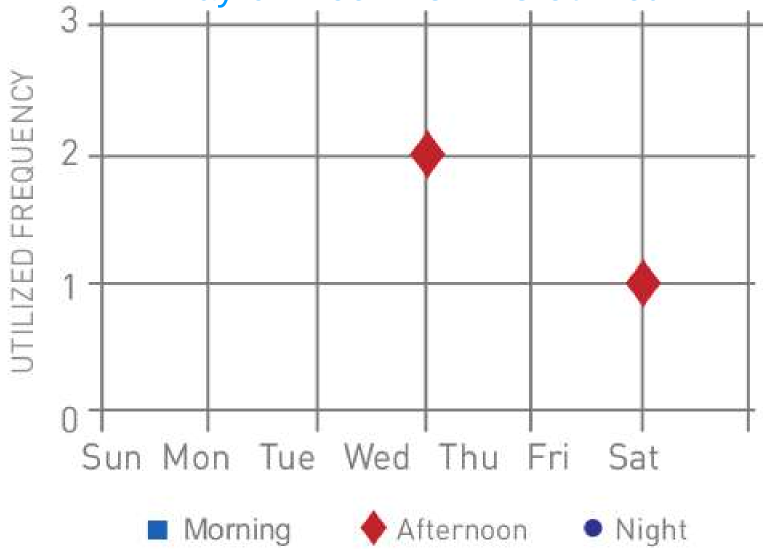


<u>Injected 3 sets of data</u>

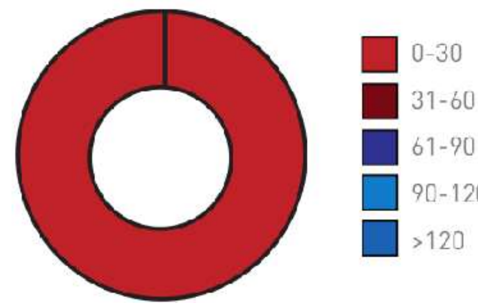
1) Notifiable disease :: Dysentery
2) Other-Communicable disease :: ADD
3) Syndrome :: Fever, Pain, RTI

TCWI Actual Usage by Health Departments

Day-of-Week TCWI is Utilized

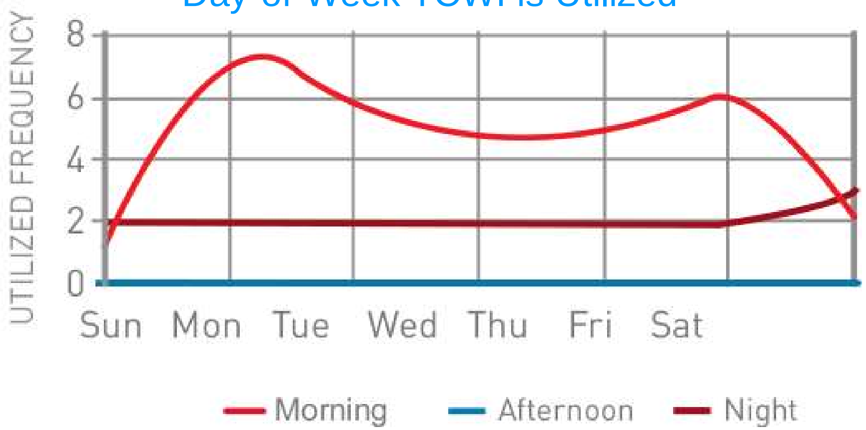


Time spent each time

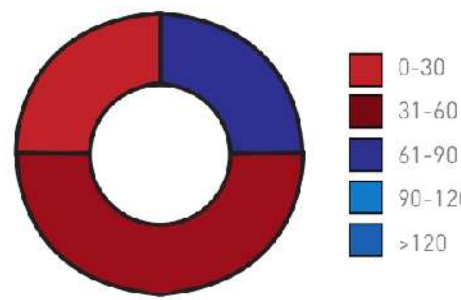


3 of 14 potential users spend less than 30 minutes each time once a week on detection analysis; remaining 9 did claim to be too busy to use TCWI

Day-of-Week TCWI is Utilized



Time spent each time

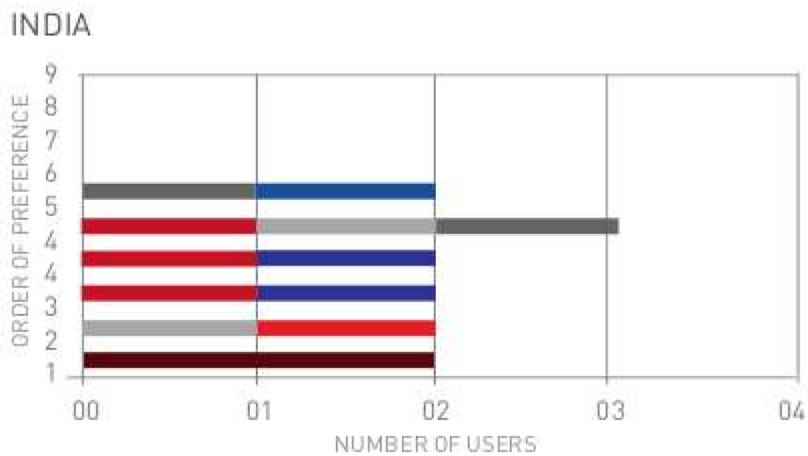


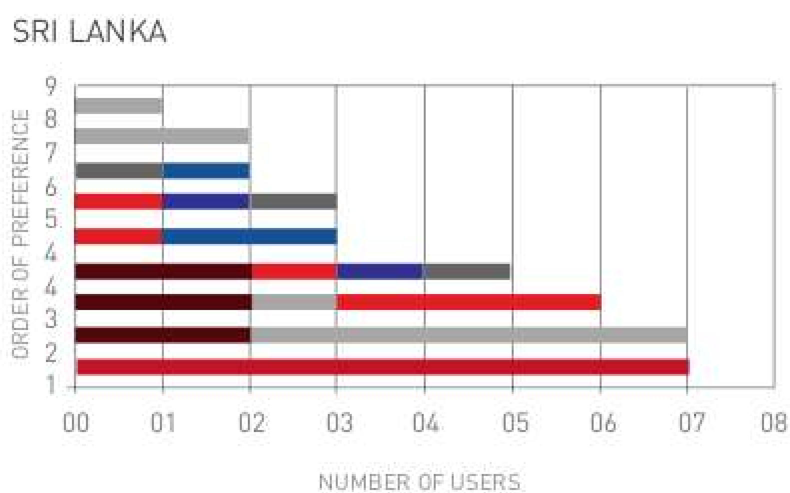
75% of the 9 Sri Lankan users spend more than 30 minutes each time every day of the week on detection analysis.

www.lirneasia.net

TCWI Preferred functions

India health officials' primary preferences are screening for fever, other-communicable diseases, and using the pivot table.





Sri Lanka health officials' primary preferences are screening the notifiable, fever, and other-communicable diseases.

Pre Screen: Temporal scan based massive screening to detect significant recent increase for cases with fever like symptoms

Pre-Screen Notifiable: Temporal scan based massive screening to detect significant recent increase for Notifiable disease

Pre-Screen Non-Communicable: Temporal scan based massive screening to detect significant recent increase for Non-Communicable disease

Spatial Scan: Algorithm to detect spatial regions with significant recent increase for a given disease

Ad Hoc Pivot Table: Pivot table function allows user to freely explore data in a multi-dimensional tabular view

Pre-Screen Other Communicable: Temporal scan based massive screening to detect significant recent increase for Communicable diseases other than notifiable diseases

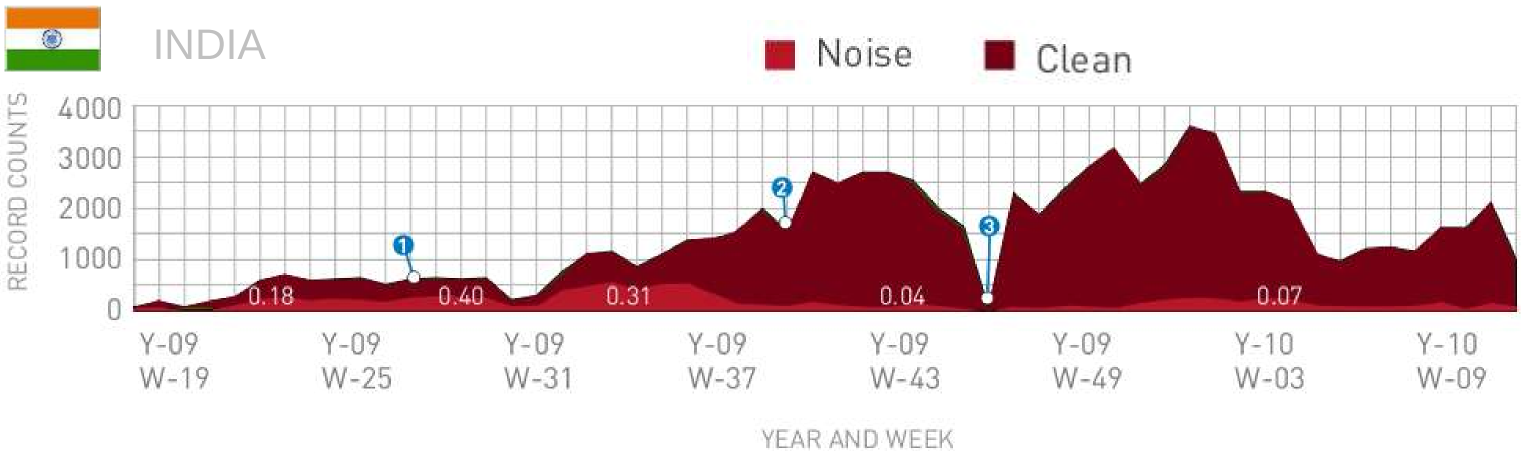
Ad-hoc Temporal Scan: Generic temporal scan method to detect any unusual temporal events comparing with a baseline

Pivot Table Weekly: Pre defined pivot table summarizing current week counts by location, disease and demographics

Other: Any other statistical analysis method available from TCWI

Quality of the digitized data

Data quality = Signal to Noise Ratio (SNR); i.e. number records with errors/records submitted



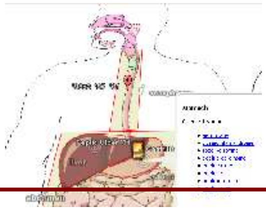
The 23% noisy data in India subsided to less than 4% after informing the consequences of false detections (SNR for sub intervals: 0.18, 0.40. 0.31. 0.04, 0.07)

Assistants in Sri Lanka with no formal health training and no affiliation to the hospitals/clinics had no incentive to correct the 45% errors (SNR for sub intervals: 0.58, 0.30, 0.53, 0.57, 0.17)

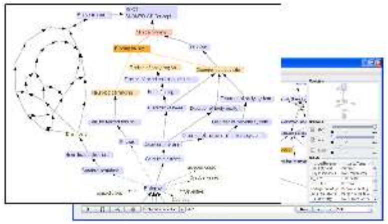
[1] Low quantities of data received from Health Sub Centers
[2] Volume of records were better after including Primary Health Centers
[3] Holiday effect: no records received
[4] Learning curve getting medical officers to adopt to the new procedures of writing the diagnosis
[5] Release of mHealthSurvey v1.3 with better predictive text

Digitizing problems that affect the categorical data

PROBLEM	EXAMPLE
Use of synonyms	goal fever =jail fever = typhus fever dementia = memory loss enteric fever = typhoid fever; encephalitis = meningitis
Inserting symbols and extra spacing between words	body ache =body-ache, body pain = body pain
Changing the order of words	muscle weakness = weakness in muscle stomach pain = pain in the stomach
Inclusion and exclusion of adjectives	'severe' memory loss vs memory loss
Using local language when terms are unknown	leg vettuthal (Tamil) = broken leg (English)
Using preposition and conjunctions between different terms	nasal stuffiness *or* sneezing over bleeding *with* abdominal pain
Long sentences	not able to identify color white and shining patches without any sense
Mistaking treatment for diagnosis	oral pils, remove catheter, vaccination
Prepopulated instructions in text boxes propagating to database"	please specify details specify symptoms
UK vs USA spelling	diarrhoea = diarrhea vomiting = vommitting
Test results as symptoms or signs	BP 140/90, BP 120/100
Singular vs plural	fit / fits , cut / cuts
Inconstancies in the verb tense	faint, fainted, fainting

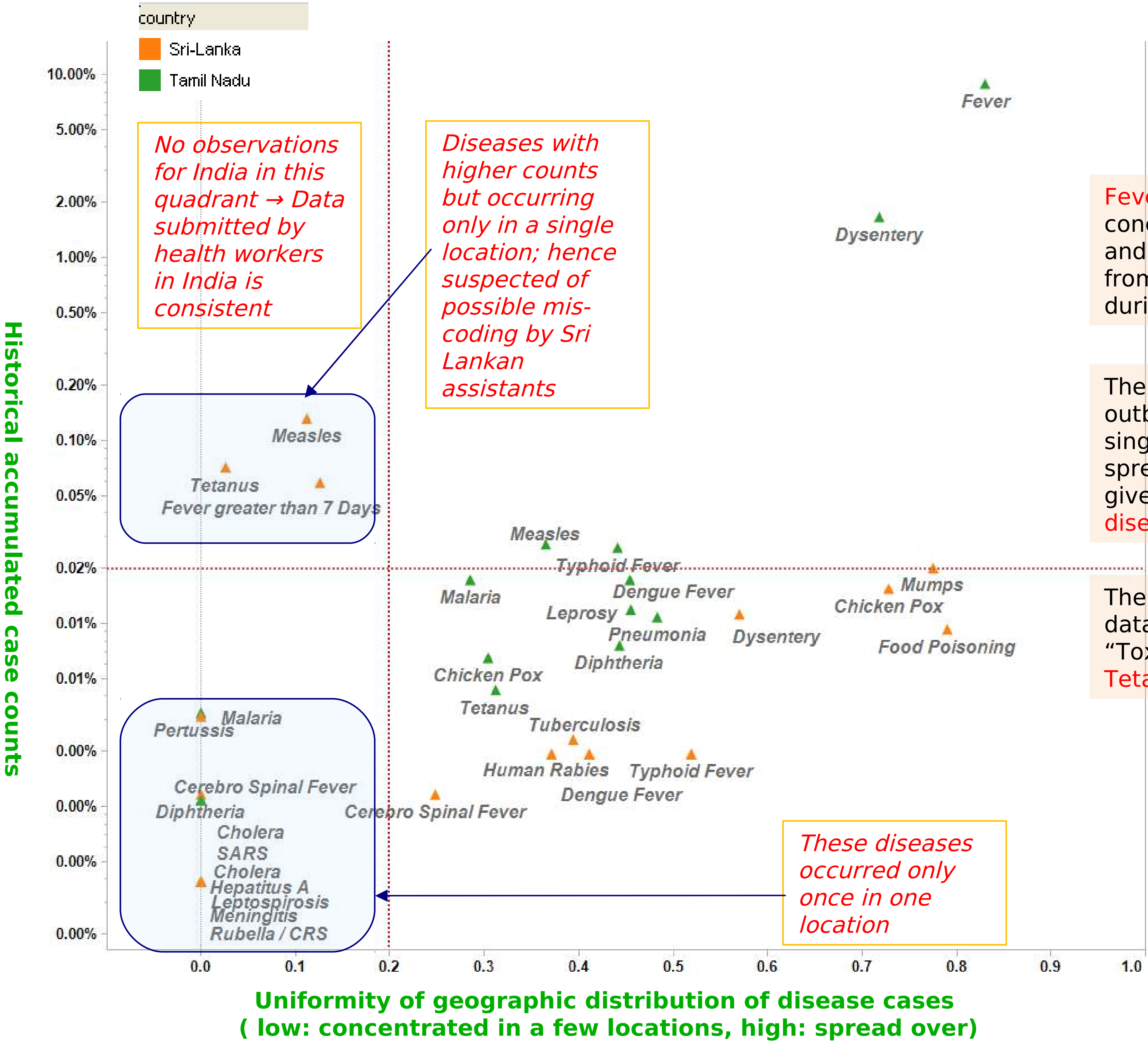


SNOMED-CT



LOINC

Observations of the data digitizing uncertainties



Fever greater than 7 days concentrated in February and March of 2010, mainly from a single location, during the non rainy season

The likelihood of a measles outbreak emerging only in a single location without spreading to other areas, given that it is a viral disease, is highly unlikely.

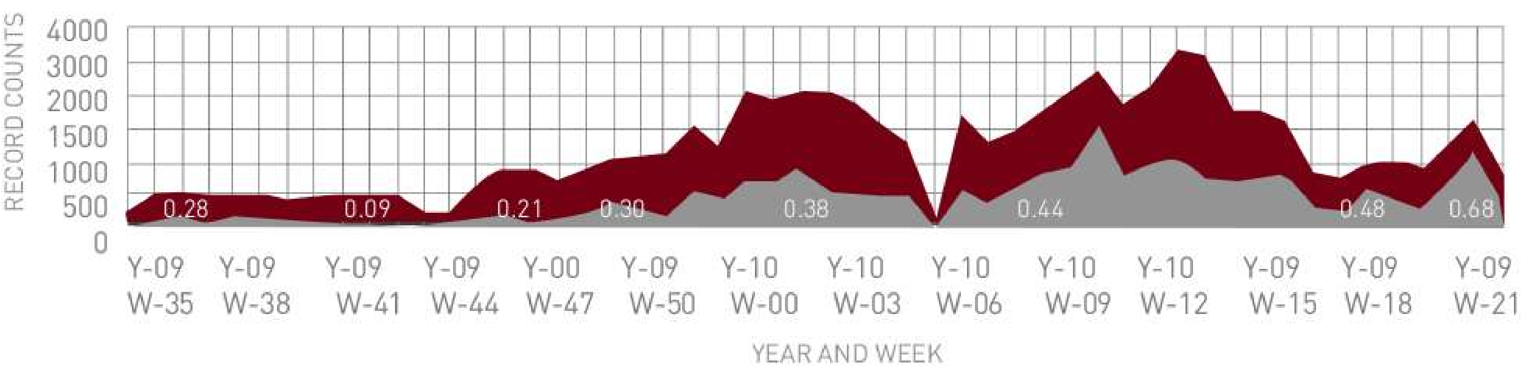
The assistant entering the data had submitted data for “Toxide vaccine” as Tetanus.

Timeliness of data submission

Timeliness = submitting the patient's record the same day as the patient visitation

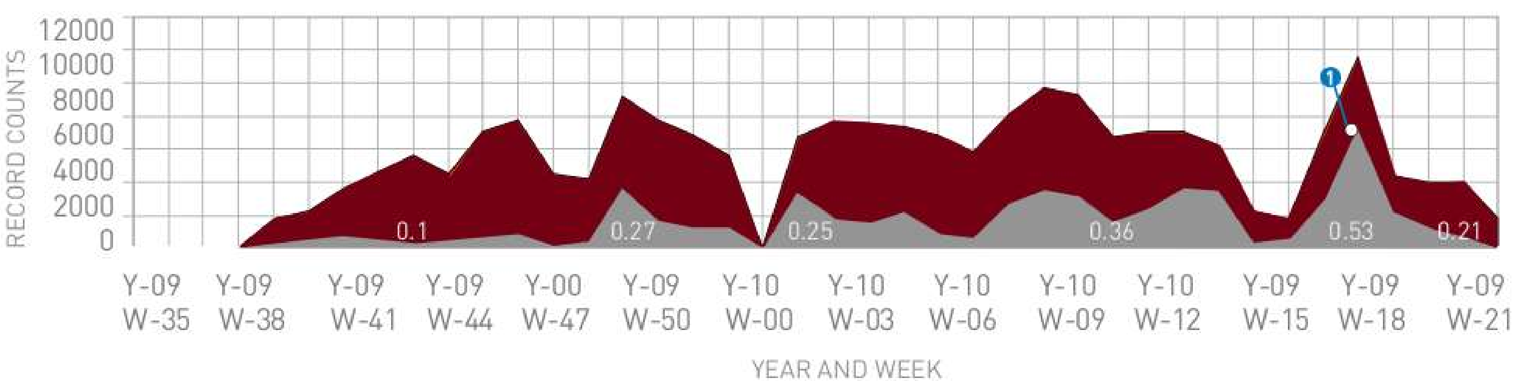
INDIA

≥ 1 day < 1 day



Finding time to complete the records without disrupting current work flow was a significant barrier for real- time data submission (sub interval delay rates 0.28, 0.09, 0.21, 0.38, 0.44, 0.48, 0.68)

SRI LANKA



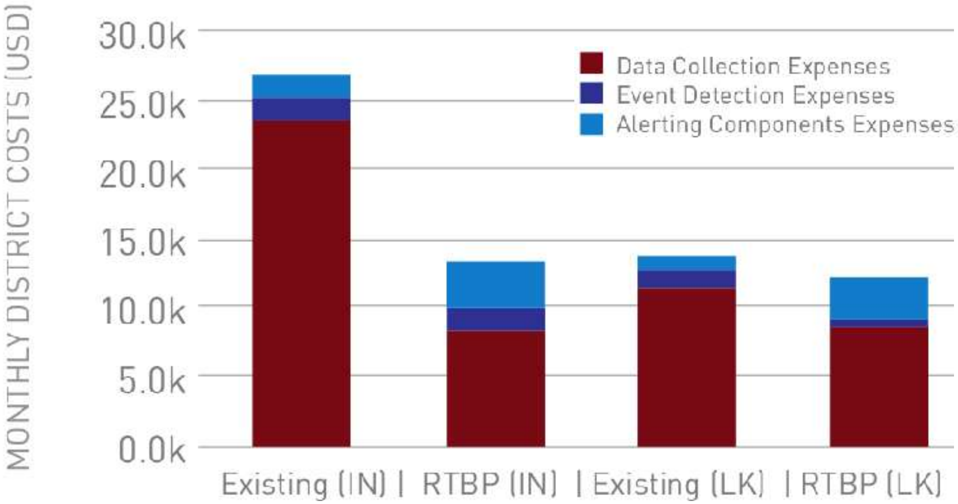
Data entry assistants have no other role besides digitizing records but see delays proportional to the patient visitation counts (sub interval delay rates: 0.10, 0.27, 0.25, 0.36, 0.53, 0.21).

[1] Users with dysfunctional phones where sharing and were sending data on the weekends or when friends phone was available for borrowing

www.lirneasia.net

Distribution of expenses calculated in terms ToC

Comparison of expenses in relation to the data collection, event detection, and alerting components

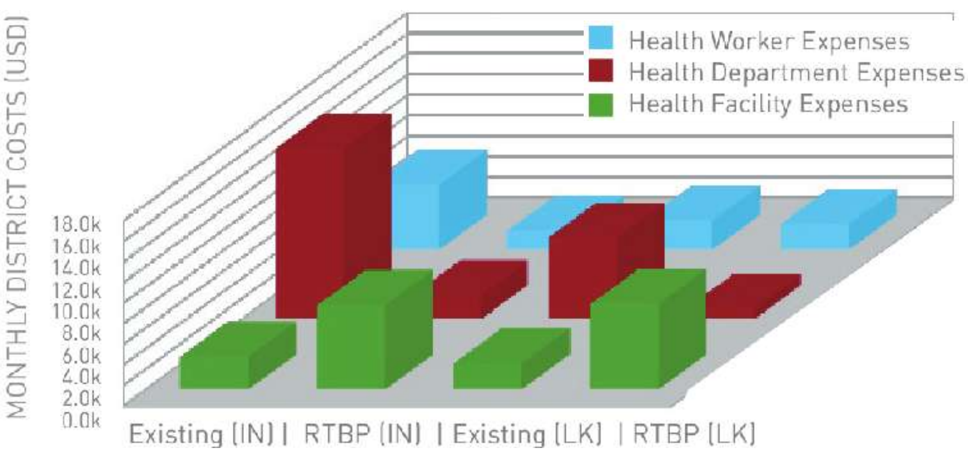


India and Sri Lanka invest very little or no resources on event detection and alerting

RTBP can reduce direct expenses, increase timeliness, and introduce detection and alerting components

India and Sri Lanka can reduce overall expenses by 50% and 30%, respectively, with ICT

Comparison of expenses in relation to the health facility, health department, and health workers

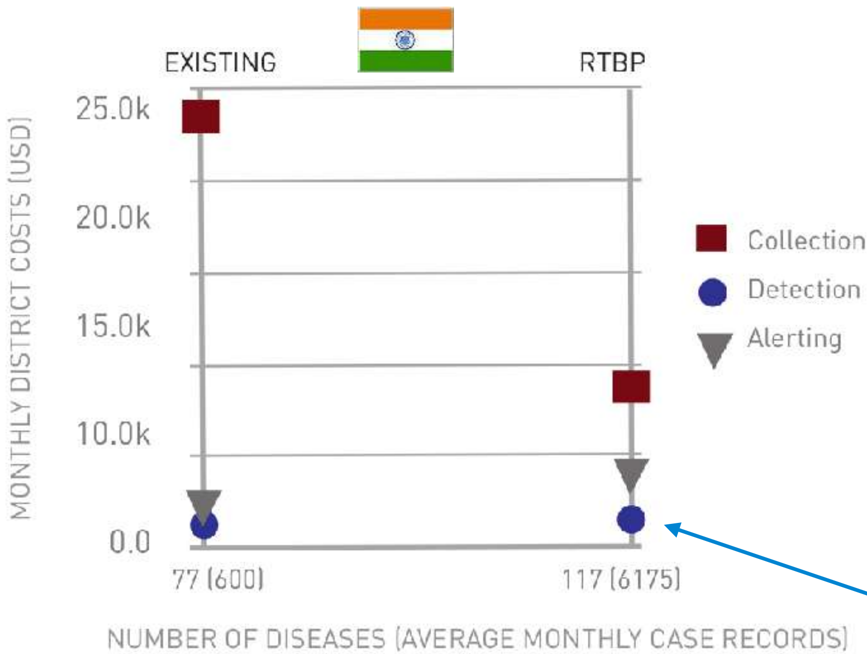


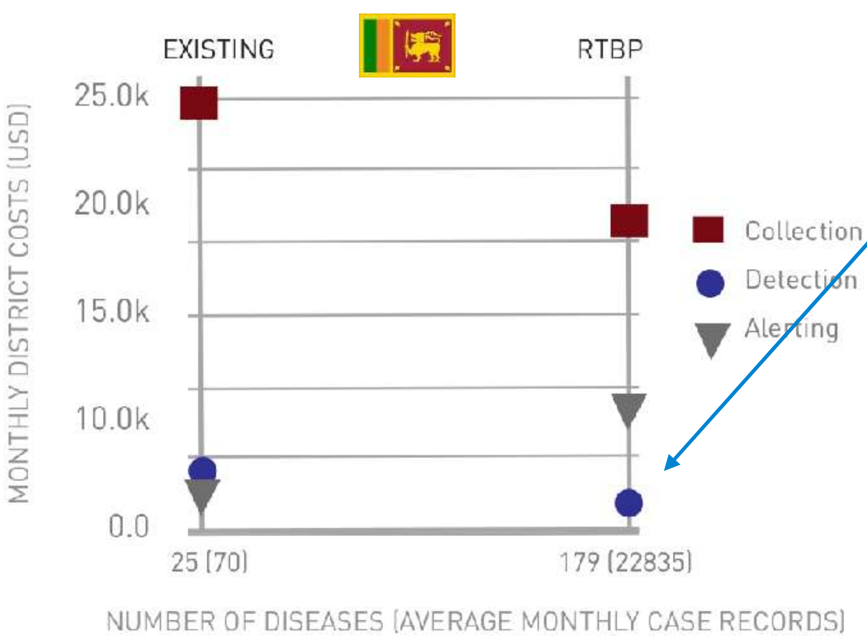
Bulk of the expenses are in health departments invested for data collection and consolidation, which can be reduced by RTBP with the introduction of mHealth at the point of care

Invest more in alerting to empower health workers with information on the state of affairs of the health in their regions

[Existing (IN) = present system in India (Integrated Disease Surveillance Program); Existing (LK) = present system in Sri Lanka (Disease Surveillance and Notification Program); RTBP (IN), RTBP (LK) = Real-Time Biosurveillance Program in India and Sri Lanka, respectively]

Incremental Cost Effectiveness Ratios (ICER)





Going from **Existing** system to **RTBP**

Collection – mHealthSurvey can reduce the expense and enhance system to collecting all data opposed to reporting a small subset of diseases once only

Detection – TCWI can introduce syndromic and disease surveillance opposed to no rapid detection analyses done at present

Alerting – not an existing practice but health workers will be better informed of the public health status in the geographic area for better response and mitigation

T-Cube Web Interface: Some Feedback

"We can use this rich and comprehensive dataset and analysis tools for our annual planning, now our planning relies on professional perception and not necessary data."

- Deputy Director Planning, Kurunegala District, Sri Lanka, Consulted (06.10.09)

"Epidemiologists want TCWI to facilitate the old ways of monitoring outbreaks based on thresholds opposed to statistical significance. For example, a single case of Malaria is regarded as an outbreak in India, which requires response actions."

- Deputy Director of Health Services, Sivaganga District, India, Consulted (19.12.09).

"It is important to monitor escalating fever cases, notifiable disease cases, and common clusters of symptoms."

- Regional Epidemiologist, Kurunegala District, Sri Lanka, consulted (19.12.09).

"Medical Officers, Nurses, Health Educators, etc, who are interested in learning of outbreaks see the benefit and are happy with TCWI detection analysis methods but the staff at the Integrated Disease Surveillance Program are not ready to accept change and want to stick to the traditional system unless state or national level Authorities mandate it."

- Senior Project Officer, RTBI, India, consulted (19.08.10).

T-Cube Web Interface: Some Feedback

"Pharmacists' perceptions are such that a separate computer should be given for detection analysis and they do not want to share their computers, which are used for medicine and birth information."

- Senior Project Officer, RTBI, India, Consulted (08.07.2010).

"RTBP's real-time biosurveillance capabilities will enhance the present day passive or non-active passive surveillance to an active surveillance system."

- Wayamba Provincial Director of Health Services, consulted (07.07.10).

"All cases can be viewed in TCWI in real-time for detecting outbreaks swiftly, which other-wise would take several days before the hospitals/clinics send the notification paper forms, by which time the patient may be dead or discharged."

- Public Health Inspector, Wariyapola, Sri Lanka, consulted (26.04.10).