THE ROLE OF LABORATORY DATA IN ‘KNOWLEDGEABLE SURVEILLANCE’

Crawford Revie (AVC / UPEI, Canada)
Fernanda Dórea (SVA, Sweden)
OUTLINE OF TALK

• ‘Knowledgeable surveillance’?
• Recent trends in AI and machine learning
• Algorithms for adding value to laboratory data
• Integration with other data/knowledge sources

“It's what you learn after you know it all that counts.”
(Harry S Truman)
knowledgeable surveillance...

epidemiological intelligence...

A.I.
Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?
“The fundamental tools of A.I. shifted from logic to probability in the late 1980s, and fundamental progress in the theory of uncertain reasoning underlies many of the recent practical advances.”

(Peter Norvig, 2011)
IBM’S RESEARCH

Deep Blue

Watson
gauge sentiment towards a cafe using ‘tweets‘:

• “Delicious muesli from the @imaginarycafe - what a great way to start the day!”
• “Greatly enjoyed the home-made lasagna at the Imaginary Café last night.”
• “It was a great disappointment to find that Imaginary Cafe stopped stocking BLTs.”
• “Had to wait in line for 45 minutes at the Imaginary Cafe today. Great, well there’s my lunch-break gone...”
IBM's Watson to take on veterinary medicine

SHANE DINGMAN - TECHNOLOGY EDITOR
The Globe and Mail
Published Tuesday, Oct. 07, 2014 7:00PM EDT
Last updated Tuesday, Oct. 07, 2014 7:00PM EDT

First International Business Machines Corp. took on Jeopardy and won with its Watson cognitive computing engine. Now, it's taking aim at animal welfare.

IBM will reveal the first 10 apps using Watson for commercial purposes on Wednesday, pushing into such areas as retail management, travel planning, medicine and computer security. Among the first apps will be Sofie, a new service made by LifeLearn Inc. of Guelph, Ont., that acts as a real-time...
WHAT HAS CHANGED?

Deep Neural Networks: A Getting Started Tutorial
06/13/2014
Deep Neural Networks are the more computationally powerful cousins to regular neural networks. Learn exactly what DNNs are and why they are the hottest topic in machine learning research.
WHAT HAS CHANGED?

Results from *ILSVRC* (ImageNet Large Scale Visual Recognition Challenge)  
Hinton et al (2012)
Eugene Wigner’s article “The Unreasonable Effectiveness of Mathematics in the Natural Sciences” examines why so much of physics can be neatly explained with simple mathematical formulas such as $f = ma$ or $e = mc^2$. Meanwhile, sciences that involve human beings rather than elementary particle behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data.

Learning from Text at Web Scale
The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The

The Unreasonable Effectiveness of Mathematics in the Natural Sciences. (Eugene Wigner, 1960)
“Meanwhile, the poor Babel fish, by effectively removing all barriers to communication between different races and cultures, has caused more and bloodier wars than anything else in the history of creation.”

(Hitch-Hikers Guide to the Galaxy, Douglas Adams)
DATA-DRIVEN GOVERNMENT

AND EVERYTHING ELSE...
BIG DATA

Volume

Variety

Velocity

19 March 2015

“The difference between something good and something great is attention to detail.”

(Charles R Swindoll)
“This is the most personal device we have ever created. It's not just with you, it's on you.”
(Tim Cook, Apple CEO)
BIG DATA - VARIETY

**Case Details**
- Farmer's Name
- Animal Identifier
- Clinician's Name
- Location or District

**Select Signs**
- □ Anaemia and Pallor (Lack of blood/pale membranes)
- □ Anorexia/Depression (Loss of appetite)
- □ Ataxia (Loss of movement balance)
- □ Constipation (Scanty/small or pellet/hard faeces)
- □ Diarrhoea
BIG DATA - VARIETY

- clinical signs
- syndromes / disease
- geo-spatial distribution
- photos of symptoms
- ID by photo (‘face’ recognition)
- body mass estimates
- temperature readings
- pen-side testing of samples
Fancy tools/techniques are useful, but they can never compensate for poor quality data...
In the context of – Syndromic Surveillance

Bio-surveillance

Total Information Management

Social media / network analysis

National Syndromic Surveillance Program (NSSP)

A collaboration among local, state, and national public health programs for timely exchange of syndromic data for situational awareness and enhanced response to hazardous events and disease outbreaks.

CDC Surveillance Strategy

A strategy for improving the Centers for Disease Control and Prevention's activities in public health surveillance
VETERINARY SYNDROMIC SURVEILLANCE

Review
Veterinary syndromic surveillance: Current initiatives and potential for development
Fernanda C. Dórea*, Javier Sanchez, Crawford W. Revie
Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, PE, C1A 4P3, Canada

ARTICLE INFO
Article history:
Received 4 February 2011
Received in revised form 5 May 2011
Accepted 8 May 2011

Keywords:
Syndrome surveillance
Veterinary surveillance
Animal health surveillance
Emerging diseases
Abnormal detection
Prospective monitoring

ABSTRACT
This paper reviews recent progress in the development of syndromic surveillance systems for veterinary medicine. Peer-reviewed and grey literature were searched in order to identify surveillance systems that explicitly address outbreak detection based on systematic monitoring of animal population data, in any phase of implementation. The review found that developments in veterinary syndromic surveillance are focused not only on animal health, but also on the use of animals as sentinels for public health, representing a further step towards One Medicine. The main sources of information are clinical data from practitioners and laboratory data, but a number of other sources are being explored. Due to limitations inherent in the way data on animal health is collected, the development of veterinary syndromic surveillance initially focused on animal health data collection strategies, analyzing historical data for their potential to support systematic monitoring, or solving problems of data classification and integration. Systems based on passive notification or data transfers are now dealing with sustainability issues. Given the ongoing barriers in availability of data, diagnostic laboratories appear to provide the most readily available data sources for syndromic surveillance in animal health. As the bottlenecks around data source availability are overcome, the next challenge is consolidating standards for data classification, promoting the integration of different animal health surveillance systems, and also the integration to public health surveillance. Moreover, the outputs of systems for systematic monitoring of animal health data must be directly connected to real-time decision support systems which are increasingly being used for disease management and control.

© 2011 Elsevier B.V. All rights reserved.

‘Infrastructure’: pharmacies, feed companies, abattoirs, markets
WHAT DO THESE DATA LOOK LIKE?

Porcine case submissions per month

Cases - pathology diagnoses
Cases - test results only

(Duizer, 2015)
WHAT DO THESE DATA LOOK LIKE?

(Duizer, 2015)
DIAGNOSES PER ‘CASE’

(Duizer, 2015)
BUT WHAT IS A SYNDROME?

<table>
<thead>
<tr>
<th>Group 12</th>
<th>FAT for Porcine parvovirus, Fetal necropsy. Fetus, fetal tissue, fetal stomach contents, female reproductive tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 11</td>
<td><em>Streptococcus suis</em> typing, Neurologic tissue, brain swab, heart, spleen, kidney</td>
</tr>
<tr>
<td>Group 10</td>
<td>Genotyping for PCV or PRRS. Serotyping for <em>E. coli</em>, PCR for Cytomegalovirus. Feces, lymph node (Not specified), pooled liver/spleen, pooled lung/spleen, tissue (Not specified), Body fluid (Not specified) Nasal swab, tonsil, serum, skin</td>
</tr>
<tr>
<td>Group 9</td>
<td>Micromineral analysis. Liver tissue.</td>
</tr>
<tr>
<td>Group 8</td>
<td>PCR for <em>M. hyosynoviae</em>. Gram stain, joint tissue, synovial membrane, joint swab</td>
</tr>
<tr>
<td>Group 7</td>
<td>PCR tests for <em>B. hyodysenteriae</em>, <em>B. pilosicoli</em>, <em>L. Intracellularis</em></td>
</tr>
<tr>
<td>Group 6</td>
<td>PCR for <em>C. perfringes</em>. Culture tube</td>
</tr>
<tr>
<td>Group 5</td>
<td>FAT for Porcine corona virus (Transmissible Gastroenteritis). Fecal smear</td>
</tr>
<tr>
<td>Group 4</td>
<td>Electron Microscopy, anaerobic culture, ELISA for <em>C difficile</em>, <em>E. coli</em> K88 serotyping. Feces (fixed), coccidial smear. Live animal submission, all intestinal tissue</td>
</tr>
<tr>
<td>Group 3</td>
<td>PCR tests for swine influenza (H3N2 &amp; H1N1), <em>M hyopneumoniae</em>, PRRS &amp; PCV. Lung tissue</td>
</tr>
<tr>
<td>Group 2</td>
<td>Necropsy. Carcass, pooled lung/tonsil</td>
</tr>
<tr>
<td>Group 1</td>
<td>Histology, aerobic culture. Fixed tissue</td>
</tr>
</tbody>
</table>

(Duizer, 2015)
ALTERNATIVE VIEW OF SYNDROMES

Based on 3 years studying laboratory data sets and expert opinion

<table>
<thead>
<tr>
<th>Abortion</th>
<th>Musculoskeletal</th>
<th>\textit{Neospora caninum}</th>
<th>Clinical pathology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circulatory</td>
<td>Nervous</td>
<td>Johnes disease</td>
<td>Toxicology</td>
</tr>
<tr>
<td>Environmental</td>
<td>Reproductive</td>
<td>BSE</td>
<td>MIC</td>
</tr>
<tr>
<td>Eyes and ears</td>
<td>Respiratory</td>
<td>BVD</td>
<td>NONSPECIFIC</td>
</tr>
<tr>
<td>GIT</td>
<td>Systemic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haematopoietic</td>
<td>Urinary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatic</td>
<td>Skin and tegument</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastitis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Dórea, 2012)
<table>
<thead>
<tr>
<th>Syndrome</th>
<th>caseID</th>
<th>Source</th>
<th>Date</th>
<th>Sample</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLV</td>
<td>2092814</td>
<td>clinical</td>
<td>2000-Jan-01</td>
<td>Serum</td>
<td>BLV</td>
</tr>
<tr>
<td>BLV</td>
<td>2092814</td>
<td>clinical</td>
<td>2000-Jan-01</td>
<td>Serum</td>
<td>BLV</td>
</tr>
<tr>
<td>BVDV</td>
<td>2092582</td>
<td>clinical</td>
<td>2000-Jan-01</td>
<td>Serum</td>
<td>BVD type 2</td>
</tr>
<tr>
<td>BVDV</td>
<td>2092607</td>
<td>clinical</td>
<td>2000-Jan-01</td>
<td>Serum</td>
<td>BVD type 2</td>
</tr>
<tr>
<td>BVDV</td>
<td>2092607</td>
<td>clinical</td>
<td>2000-Jan-01</td>
<td>Serum</td>
<td>BVD type 2</td>
</tr>
<tr>
<td>Hepatic</td>
<td>2000032</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Liver</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Respiratory</td>
<td>2000032</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Lung</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Haematopoietic</td>
<td>2000032</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Spleen</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Nonspecific</td>
<td>2000128</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Tissue___Fixed</td>
<td>Histopathology</td>
</tr>
<tr>
<td>Nonspecific</td>
<td>2000055</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Tissue___Fixed</td>
<td>Histopathology</td>
</tr>
<tr>
<td>BVDV</td>
<td>2000032</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Tissue___Fixed</td>
<td>BVD type 1</td>
</tr>
<tr>
<td>Nervous</td>
<td>2008045</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Brain</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Nervous</td>
<td>2008001</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>brain</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Abortion</td>
<td>2015705</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Fetal_Tissue</td>
<td>Bacterial Culture</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>2015789</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>bone</td>
<td>Clostridia</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>2015789</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>joint</td>
<td>Clostridia</td>
</tr>
<tr>
<td>Respiratory</td>
<td>2015719</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Lung</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>GIT</td>
<td>2015705</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>rumen</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>GIT</td>
<td>2015705</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>rumen</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>Respiratory</td>
<td>2015705</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Bronchoalveolar_Lavage</td>
<td>Bacterial culture</td>
</tr>
<tr>
<td>GIT</td>
<td>2024811</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>large Int</td>
<td>Clostridium perfringens</td>
</tr>
<tr>
<td>Respiratory</td>
<td>2024699</td>
<td>clinical</td>
<td>2000-Jan-02</td>
<td>Lung</td>
<td>Virus isolation</td>
</tr>
</tbody>
</table>
CAN WE AUTOMATICALLY CLASSIFY?

Direct mapping:
Based on keywords – 25% of all records

Text mining algorithms:
Dictionary of relevant words
Direct relationships
Co-occurrences
Precedence
**CAN WE AUTOMATICALLY CLASSIFY?**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Recall</th>
<th>Precision</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manually modified rules</td>
<td>0.994</td>
<td>1.000</td>
<td>0.997</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.983</td>
<td>0.939</td>
<td>0.955</td>
</tr>
<tr>
<td>Rule Induction</td>
<td>0.626</td>
<td>0.793</td>
<td>0.677</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>0.290</td>
<td>0.416</td>
<td>0.311</td>
</tr>
</tbody>
</table>

*(Dórea et al, 2012)*
Records classification

Data extraction

Classification of records into syndromes

Case counts for each syndrome

Retrospective analysis

Removal of aberrations and excessive noise + Characterization of temporal effects

Prospective analysis

Use historical data as a baseline to train algorithms for prospective detection of aberrations

Alarm detection + Automated correction of baseline

Real-time monitoring

Alarms of various magnitudes
The majority of submissions were associated with a specific syndrome

(Dórea et al, 2013)
VARIATION IN SYNDROME FREQUENCY

![Graph showing variation in syndrome frequency over years 2008 to 2011. The graph includes two subplots: one for Mastitis and one for Reproductive. The Mastitis graph shows a range from 0 to 30 on the y-axis, while the Reproductive graph shows a range from 0 to 5 on the y-axis.]
BASELINE - RETROSPECTIVE ANALYSIS

Noise versus past outbreaks
Small counts per day
Temporal effects
BASELINE - RETROSPECTIVE ANALYSIS

Day of the week (DOW)
Monthly variation (explicit vs Sine fct.?)
Poisson regression model with DOW and monthly effects for BLV at daily submission.

(Dórea et al, 2012)
PROSPECTIVE ANALYSIS AS A MEANS TO ABERRATION DETECTION

Which algorithms?

- Control charts
  - Shewhart
  - Cumulative sums (CUSUM)
  - Exponentially Weighted Moving Averages (EWMA)

- Regression models + Control chart
  - Poisson, Negative Binomial
  - Differencing + Control charts

- Holt-Winters exponential smoothing
BUT WHAT MIGHT AN OUTBREAK LOOK LIKE?
The graphs illustrate the sensitivity of various statistical models (Shewhart, CUSUM, EWMA, Holt-Winters) in detecting outbreaks of different conditions (Mastitis, BLV, Respiratory) under different outbreak signal shapes (Spike, Flat, Linear, Exponential, LogNormal).

- **Shewhart**: Sensitivity is plotted against false alarms for each condition.
- **CUSUM**: Similar to Shewhart, but with slightly different sensitivity curves.
- **EWMA**: Shows a more rapid increase in sensitivity as false alarms increase.
- **Holt-Winters**: Exhibits a slower increase in sensitivity compared to EWMA.

The graphs highlight the trade-offs between sensitivity and false alarms for each condition, with the outbreak signal shape influencing the performance of the models.
Syndromic surveillance using veterinary laboratory data: data pre-processing and algorithm performance evaluation

Fernanda C. Dórea¹, Beverly J. McEwen², W. Bruce McNab³, Crawford W. Revie¹ and Javier Sanchez¹

¹Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada C1A 4P3
²Animal Health Laboratory, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada N1H 6R8
³Ontario Ministry of Agriculture Food and Rural Affairs, Guelph, Ontario, Canada N1G 4Y2

(Dórea et al, 2013)
IMPLEMENTATION: ABERRATION CORRECTION AND INTERPRETATION

Prediction interval (upper interval)

Current data point
Alarm detected

NO Alarm detected

Guard-band or range

baseline
IMPLEMENTATION: ABERRATION CORRECTION AND INTERPRETATION

Holt-Winters exponential smoothing

Weekly Differencing + EWMA

Weekly Differencing + Shewhart
Records classification

Removal of aberrations and excessive noise + Characterization of temporal effects

Case counts for each syndrome

Use historical data as a baseline to train algorithms for prospective detection of aberrations

Prospective analysis

Alarm detection + Automated correction of baseline

Real-time monitoring

Alarms of various magnitudes
Development of an *R* package for syndromic surveillance: translating research into tools for the veterinary epidemiologist

Fernanda C. Dórea¹, Stefan Widgrén¹, Crawford W. Revie², Ann Lindberg¹

¹- Swedish Zoonosis Centre. Department of Disease Control and Epidemiology. National Veterinary Institute (SVA), Sweden.

**Abstract**

We describe the development of an *R* package for veterinary syndromic surveillance. Initial development has focused on the use of health data already classified into syndromic groups. Algorithms are available to conduct retrospective analysis of data and create outbreak-free baselines that can serve as training sets. In addition, a range of aberration detection algorithms have been proven to efficiently detect outbreak signals in animal health data streams of various magnitudes (median number of observations per day from one to hundreds) and containing various temporal effects. Functions are available to easily develop an html interface for such systems and set up automated emails in case of alarms.

**Discussion**

*R* packages are available which support specific data analysis algorithms, but their application requires experienced users capable of incorporating these algorithms into the larger framework of a syndromic surveillance system. We propose a package in which various statistical analyses already tested for use in animal data streams will be available, coupled with data management capabilities and output visualization. The main innovation offered by the package will be the ability to manage data streams, analyses, alarms and the user interface in a continuous flow; all contained in a single, open source package. This will facilitate the implementation of syndromic surveillance systems by veterinary epidemiologists.
Retrospective data analysis

Automated summary:

Creation of an outbreak-free baseline

`clean_baseline()`
Regression using GLM

`clean_baseline_perc()`
Moving percentiles

Syndromic data

The S4 class center object

`raw_to_syndromic()`

Syndromic

Observed (t x S)
Dates (t x 9)
Baseline (t x S)
Alarms (t x S x A)
UCL (t x S x A)
LCL (t x S x A)

Provided at creation
Result of analysis

t = # of time points
S = # of syndromic groups
A = # of algorithms used
# Daily report of syndromes in Swine 2013-11-20

## Syndromes monitored DAILY - Number of SUBMISSIONS

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Submissions</th>
<th>Main Alarm Today</th>
<th>Secondary alarms</th>
<th>Previous Days History</th>
</tr>
</thead>
<tbody>
<tr>
<td>Musculoskeletal</td>
<td>132</td>
<td></td>
<td></td>
<td>245 114 95 85</td>
</tr>
<tr>
<td>Respiratory</td>
<td>37</td>
<td></td>
<td></td>
<td>2 2 0 4</td>
</tr>
<tr>
<td>Fallen Stock</td>
<td>0</td>
<td></td>
<td></td>
<td>1 1 1 0</td>
</tr>
<tr>
<td>GIT-Bacterial tests</td>
<td>0</td>
<td></td>
<td></td>
<td>2 0 0 0</td>
</tr>
</tbody>
</table>

## Syndromes monitored WEEKLY - Number of SUBMISSIONS

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Submissions</th>
<th>Main Alarm This Week</th>
<th>Previous Weeks History</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIT</td>
<td>2</td>
<td></td>
<td>W-1 4 W-2 1 W-3 1 W-4 0</td>
</tr>
<tr>
<td>Systemic</td>
<td>4</td>
<td></td>
<td>W-1 0 W-2 1 W-3 4 W-4 0</td>
</tr>
<tr>
<td>AntimResistance</td>
<td>6</td>
<td></td>
<td>W-1 6 W-2 13 W-3 6 W-4 1</td>
</tr>
</tbody>
</table>
SOFTWARE TOOLS AND RESOURCES

Animal health

Animal health surveillance

Veterinary Syndromic surveillance

Monitoring risk factors (hazard distribution, Vectors, Geographical and Temporal risks)

Monitoring production indicators

AHSO
Animal Health Surveillance Ontology (AHSO)

"Microbial growth on vaginal swab" → Syndrome = Reproductive

"Histological examination on sample of uterus" → Syndrome = Reproductive

"Serological test for Brucellosis" → Syndrome = Reproductive

"PCR test for Brucellosis" → Syndrome = Reproductive

"Salmonella serotyping on vaginal swab" → Syndrome = Reproductive

Consensus and standards for recording data?

We don’t need ‘smarter’ data

But to get the right data to the right place

So that smart applications can do their work
An **ontology** defines a common vocabulary for those who need to share information in a domain. It includes *machine-interpretable* definitions of basic concepts in the domain and relations among them.

- **Purpose:**
  - to share common understanding of the structure of information among people or software agents
  - to enable **reuse** of domain knowledge
  - to make domain **assumptions** explicit
  - to separate **domain knowledge** from the **operational knowledge**
<table>
<thead>
<tr>
<th>Organ System</th>
<th>Clinical Signs</th>
<th>Threats</th>
<th>Lab Test Ordered</th>
<th>Test Results</th>
<th>Abattoir finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproductive</td>
<td>Reproductive</td>
<td>Non-Infectious Threats</td>
<td>Unspecific Etiology</td>
<td>Serology</td>
<td>Ante-Mortem Inspection</td>
</tr>
<tr>
<td>Female Genitalia</td>
<td>Abortion</td>
<td>Infectious Disorders</td>
<td>Bacterial Growth</td>
<td>Agent Identification</td>
<td>Post-Mortem Inspection</td>
</tr>
<tr>
<td>Vulva</td>
<td>Conception</td>
<td>Metabolic Disorders</td>
<td>Histology</td>
<td>Serological Result</td>
<td>Partial Condemnations</td>
</tr>
<tr>
<td>Vagina</td>
<td>failure</td>
<td>Infectious agents</td>
<td>Serological Tests</td>
<td></td>
<td>Total condemnations</td>
</tr>
<tr>
<td>Female Internal Reproduct. Organs</td>
<td>OTHER SYNDROMES..</td>
<td>Brucella sp</td>
<td>Brucellosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uterus</td>
<td></td>
<td>Brucella abortus</td>
<td>BVDV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovary</td>
<td></td>
<td>Brucella suis</td>
<td>Pathogen Identification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male External Genitalia</td>
<td></td>
<td>Brucella canis</td>
<td>Brucellosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male Internal Reproductive Organs</td>
<td></td>
<td>Neospora caninum</td>
<td>BVDV</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IBR</td>
<td>Pathogen Identification</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
But on the rules/relationships...

Organ System
- Reproductive
  - Female Genitalia
  - Vulva
  - Vagina
  - Female Internal Reproductive Organs
    - Uterus
    - Ovary
  - Male External Genitalia
  - Male Internal Reproductive Organs
- OTHER SYSTEMS

Clinical Signs
- Reproductive
  - Abortion
  - Conception failure
- OTHER SYNDROMES...

Threats
- Non-Infectious Threats
- Metabolic Disorders
- Infectious agents
  - Brucella sp
    - Brucella abortus
    - Brucella suis
    - Brucella canis
  - Neospora caninum
  - BVD
  - IBR

LabTestOrdered
- Unspecific Etiology
- Bacterial Growth
- Histology
- Serological Tests
  - Brucellosis
  - BVDV
- Pathogen Identification
  - Brucellosis
  - BVDV
  - Salmonella

TestResults
- Serology
- Agent Identification
- Serological Result

Abattoir finding
- Ante-Mortem Inspection
- Post-Mortem Inspection
- Partial Condemnations
- Total condemnations
Detecting influenza epidemics using search engine data


Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, S. Smolinski & Larry Brilliant

Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA.
Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA.
ILI percentages estimated by the model (black) and provided by the CDC (red) showing data available at four points in the 2007-2008 influenza season.
Big data: are we making a big mistake?

By Tim Harford

Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media.
The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,1,2* Ryan Kennedy,1,2,4 Gary King,1 Alessandro Vespignani1,5,6

In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (3, 4), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can predict x has become commonplace (5–7) and is often put in sharp contrast with traditional methods and biomarker surveillance and construct validity and reliability (8, 9). When the algorithm in 2009, and this model has run ever since, with a few changes announced in October 2013 (10, 15).

Although not widely reported until 2013, the new GFT has been persistently overestimating flu prevalence for a much longer time. GFT also missed by a very large margin in the 2011–2012 flu season and has missed high for 105 out of 108 weeks starting with August 2011 (see the graph). These errors are not randomly distributed. For example, last week's errors predict this week's errors (temporal autocorrelation), and the direction and magnitude of error varies with the time of year (seasonality). These patterns mean that GFT overlooks considerable information that could be extracted by traditional statistical methods.
Can the use of Lab Data be Effective...? (Even ‘unreasonably’ so!)

- Large collections of data will continue to become increasingly important
  - Semantic Web / Linked Open Data

- Appropriate methods will be refined:
  - novel Regression and Classification approaches
  - Cluster detection / Visualisation
  - Time series / Aberration detection
  - Structured discovery in Bayesian Networks
ACKNOWLEDGEMENTS

• Fernanda Dórea (SVA)
• Javier Sanchez (UPEI)
• Ann Lindberg (SVA)
• Beverly McEwen (AHL)
• Bruce McNab (OMAFRA)
• Ann Muckle (UPEI)