



# Assessing the Performance of American Chief Complaint Classifiers on Victorian Syndromic Surveillance Data

Bahadorreza Ofoghi  
Karin Verspoor

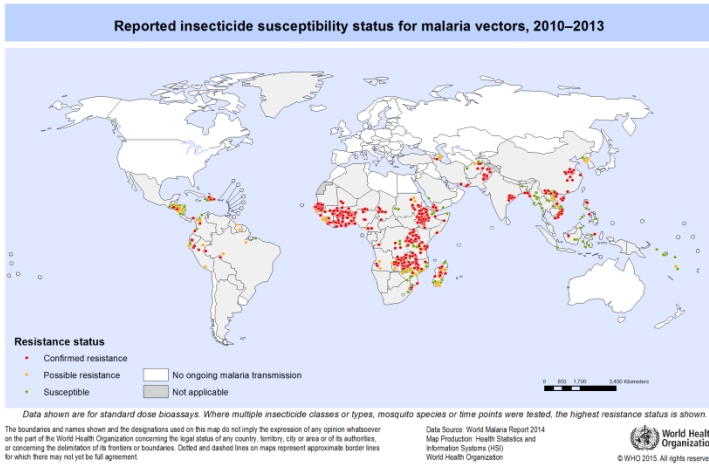
Department of Computing and Information Systems  
The University of Melbourne

October 2015



## Outline

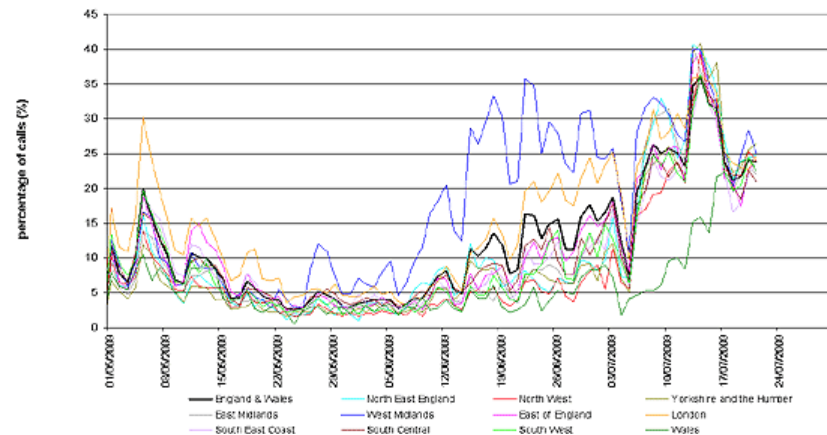
Syndromic surveillance  
Chief complaint/analysis  
RODS coders  
SynSurv data set  
RODS coders + SynSurv  
Lessons learnt



Covers:

- Early warning systems
- Outbreak detection systems
- And more...

- ...*timely analysis, detection, and investigation of significant health alerts*





*The primary symptom that a patient states as the reason for seeking medical care [medical dictionary]*

- ... captured by a nurse at an ED
- ... short, terse, abbreviated, ungrammatical
- ... ISDS recommended data

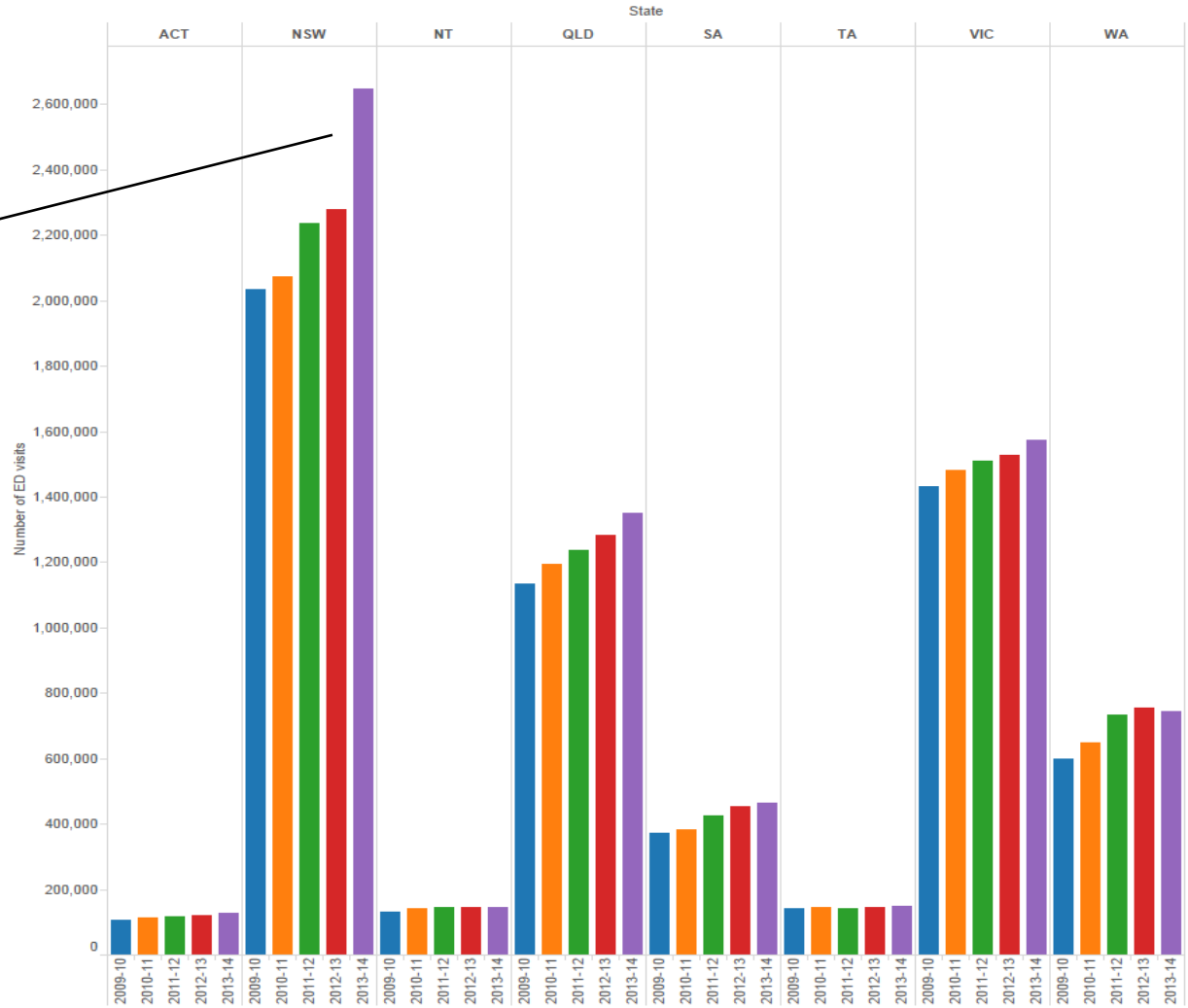
FEVER SINCE D/C LAST WEEK. NOW C/O CHEST PRESSURE, SOB. ? DX VIRAL ILLNESS AROUND HEART.





# Chief Complaint | distribution per AU state/year

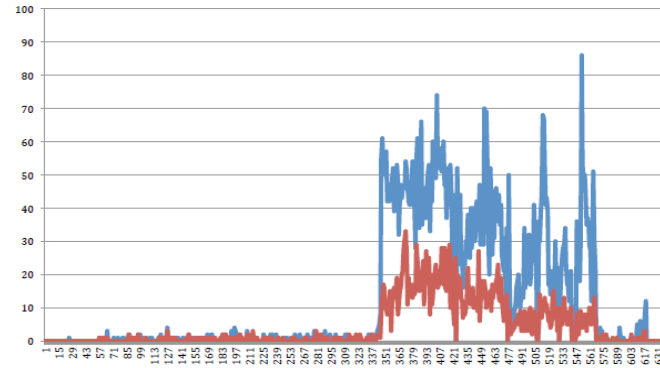
2,646,415  
ED visits in NSW  
in the FY 2013-14



Data:  
AIHW, 2013-14



# Chief Complaint Analysis



Classify and count

Code and count

Lexical analysis

FEVER SINCE D/C LAST WEEK. NOW C/O CHEST PRESSURE, SOB. ? DX VIRAL  
ILLNESS AROUND HEART.



- Real-time Outbreak and Disease Surveillance System
  - At RODS lab, University of Pittsburgh
  - Since 1999
  - Public health surveillance system
  - Collects/analyses disease data
  - Used in inter/national departments





- Implements
  - Symptom Coder (SyCo)
    - Naïve Bayes
    - Word-level
    - Two-layer: CC to symptom to syndromic group
    - Needs training
  - Complaint Coder (CoCo)
    - Naïve Bayes
    - Word-level
    - Can be used with/without training





Can we use RODS on Australian syndromic surveillance data?



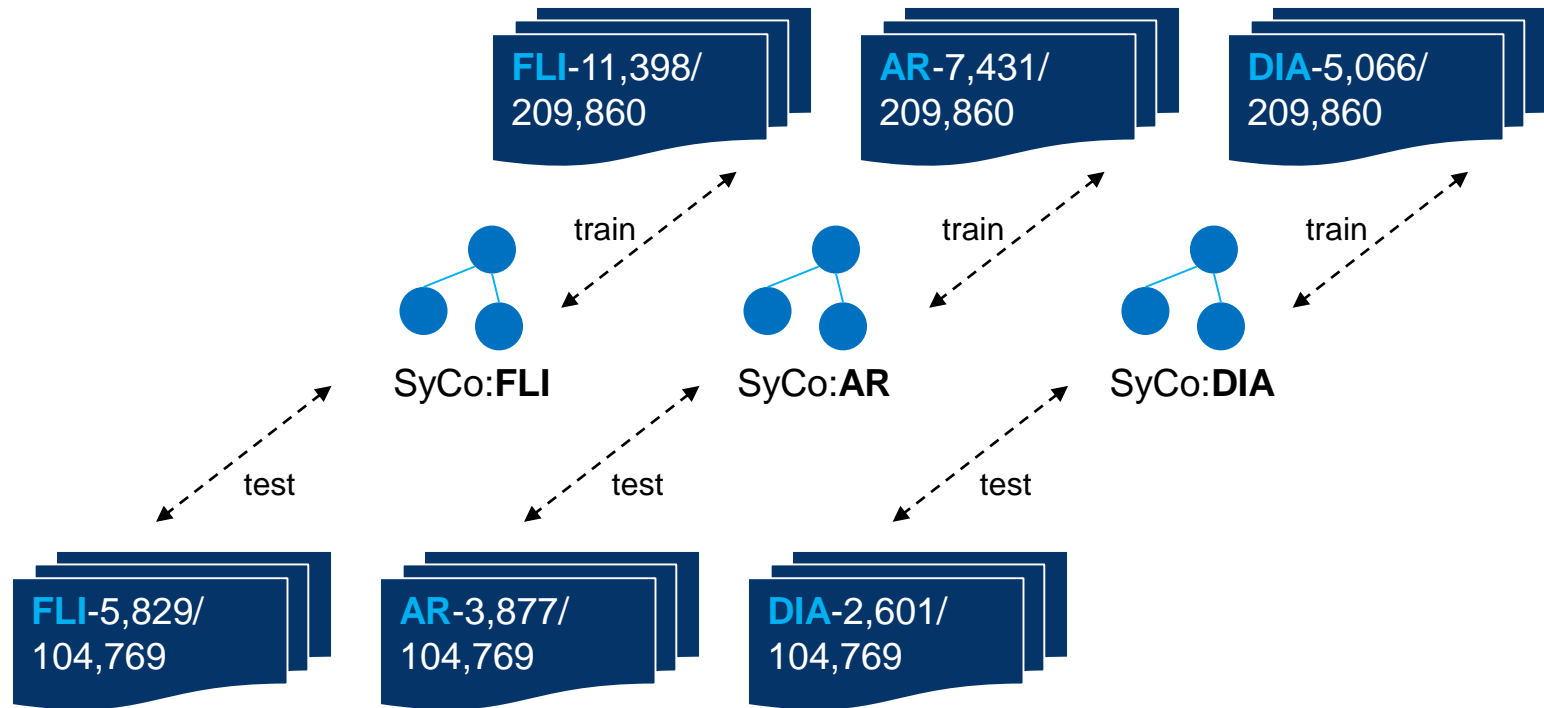
- EDs at:
  - Royal Melbourne Hospital
  - The Alfred Hospital
  - Mostly, 2000 to 2009

**Table 1.** The distribution of the different syndromic groups in the partial SynSurv data set used for our analysis

Data set	Syndromic group	#CCs
Training	Flu like illness	11,398
	Acute respiratory	7,431
	Diarrhoea	5,066
	Other	185,965
	Total:	209,860
Testing	Flu like illness	5,829
	Acute respiratory	3,877
	Diarrhoea	2,601
	Other	92,462
	Total:	104,769



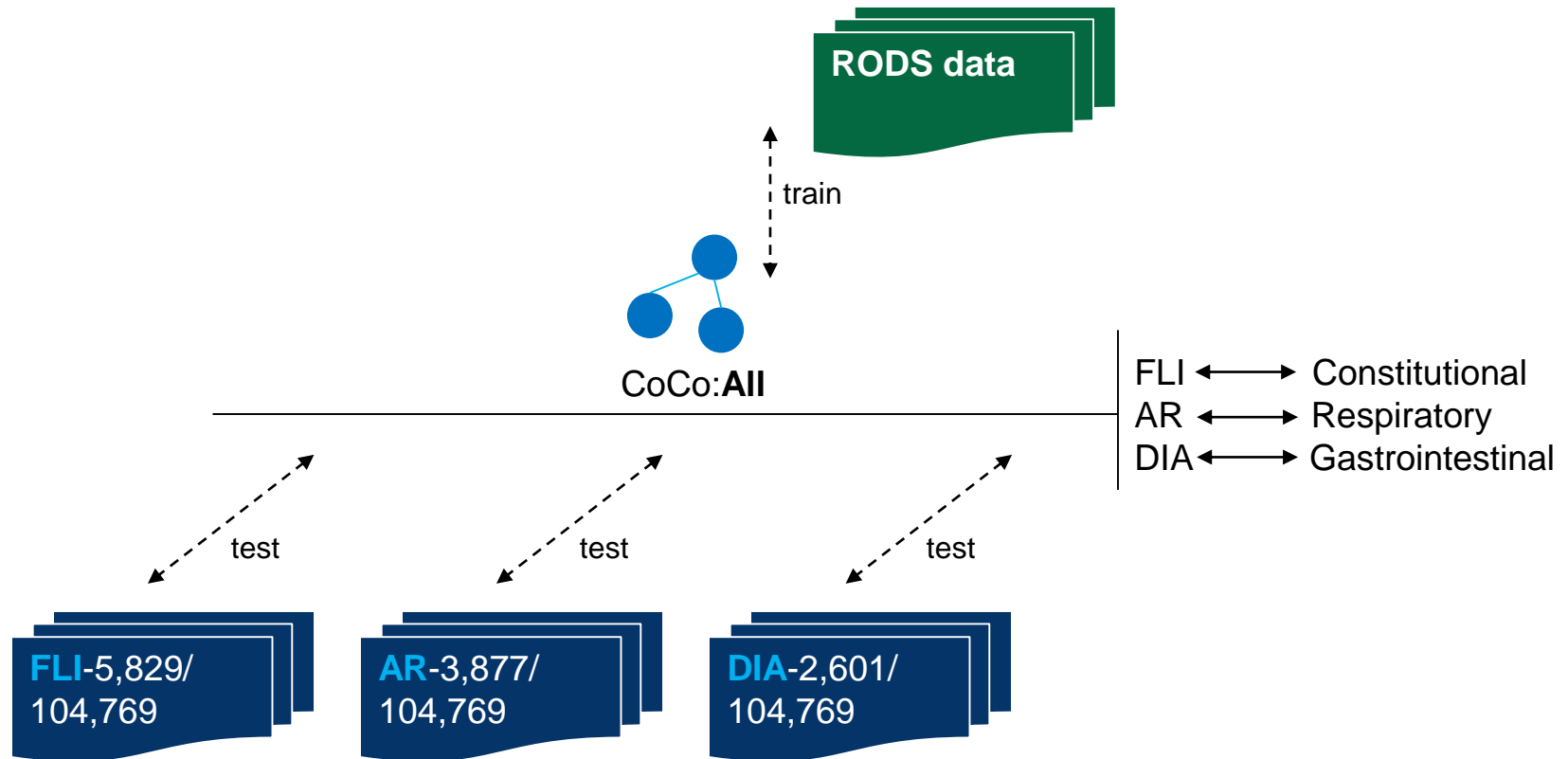
- SyCo



**FLI:** Flu Like Illness  
**AR:** Acute Respiratory  
**DIA:** Diarrhoea



- CoCo



**FLI:** Flu Like Illness  
**AR:** Acute Respiratory  
**DIA:** Diarrhoea



- Baseline systems
  - All+
    - Every test instance labelled +
  - Rand: 50-50
    - 50% chance +
    - 50% chance -
    - 10 runs



**Table 2.** The results of the classification methods with respect to each syndromic group.

Method	Syndrome	#Pos.	#TPs	#FPs	Prec.	Rec.	F1
CoCo	Flu like illness	5829	910	3998	0.1854	0.1561	0.1695
	Acute respiratory	3877	1038	1843	0.3603	0.2677	0.3072
	Diarrhoea	2601	970	4219	0.1869	0.3729	0.2490
SyCo	Flu like illness	5829	3135	5542	0.3613	0.5378	0.4322
	Acute respiratory	3877	1932	5845	0.2484	0.4983	0.3316
	Diarrhoea	2601	1084	3665	0.2283	0.4168	0.2950
All+	Flu like illness	5829	5829	98941	0.0556	1.0000	0.1054
	Acute respiratory	3877	3877	100893	0.0370	1.0000	0.0714
	Diarrhoea	2601	2601	102169	0.0248	1.0000	0.0484
Rand: 50-50	Flu like illness	5829	2922.9	49466.5	0.0558	0.5014	0.1004
	Acute respiratory	3877	1941.7	50464.8	0.0370	0.5008	0.0690
	Diarrhoea	2601	1294.7	50988.2	0.0248	0.4978	0.0472

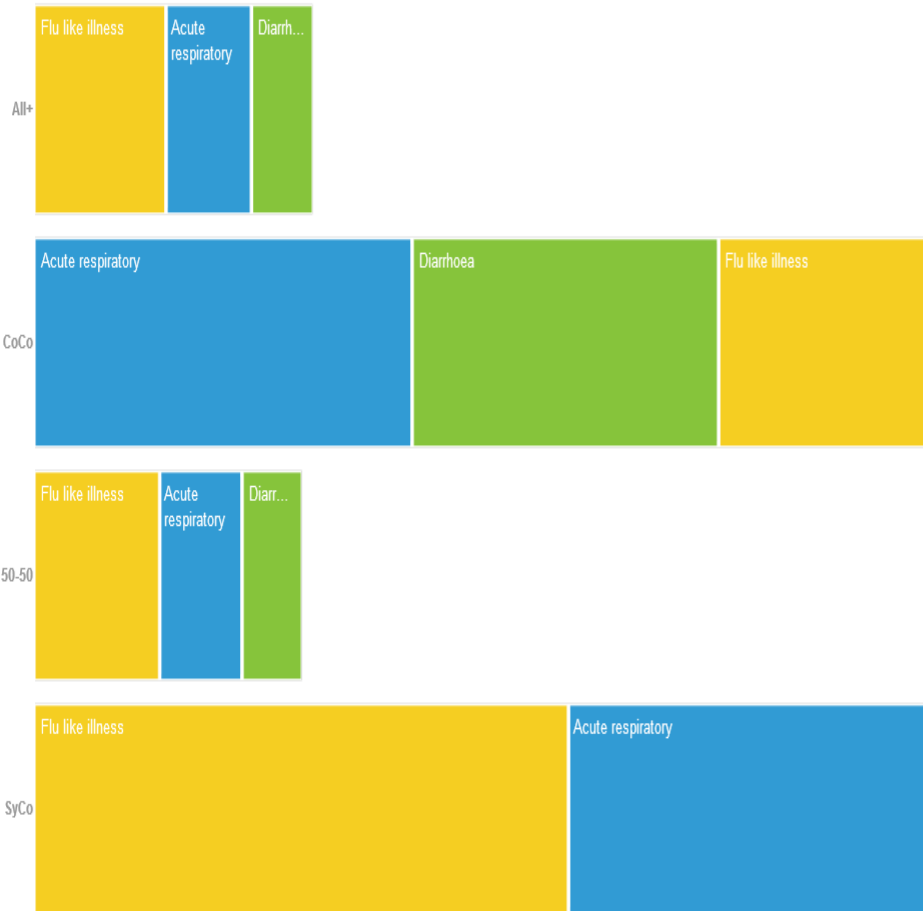
Top Method by Rec. **0.12** is the lowest Prec. for Method All+

F1 by #FPs **2** is the highest Rec. for Syndrome Diarrhoea

Top #Pos. by F1 **0.68** is the highest Prec. for Syndrome Acute respiratory

**0.43** is the highest F1 for #TPs 3135

What is the breakdown of F1 by Method and Syndrome ?



Columns

F1

Syndrome - #TPs

- Acute respiratory
- Diarrhoea
- Flu like illness



- SyCo vs. CoCo
  - Coverage
    - Dissimilar
    - Complementary
    - Next...?

**Table 3.** The agreement analysis between the classification results of SyCo and CoCo per syndromic group.

Syndromic group	Kappa	<i>p</i> -value
Flu like illness	0.1349	<.0001
Acute respiratory	0.2225	<.0001
Diarrhoea	0.2862	<.0001

<0	Poor
0 – 0.2	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

Landis & Koch, 1977





- This work...
  - RODS coders work on the scale of SynSurv
  - Outperforming trivial BLs
  - Two coders work dissimilarly
  - Adaptation to AU data
    - SyCo: n/a
    - CoCo: syndromic groups
- Next...
  - Ensembling RODS coders
  - Other classifiers
    - Our work to be published soon!



**Thank  
you**

Bahadorreza Ofoghi

[bahadorreza.ofoghi@unimelb.edu.au](mailto:bahadorreza.ofoghi@unimelb.edu.au)