TRANSFoRm

Data Mining Primary Care Data as part of the TRANSFoRm Project

Przemyslaw Kazienko, Tomasz Kajdanowicz - Wroclaw University of Technology, Poland

Roxana Danger Mercaderes, Vasa Curcin – Imperial College London

Jean-Karl Soler – Mediterranean Institute for Primary Care, Malta

Derek Corrigan – Royal College of Surgeons in Ireland

Brendan Delaney – Kings College London

EGPRN – Malta – October 2013

This project is partially funded by the European Commission under the 7th Framework Programme. Grant Agreement No. 247787 Translational Research and Patient Safety in Europe (TRANSFoRm)

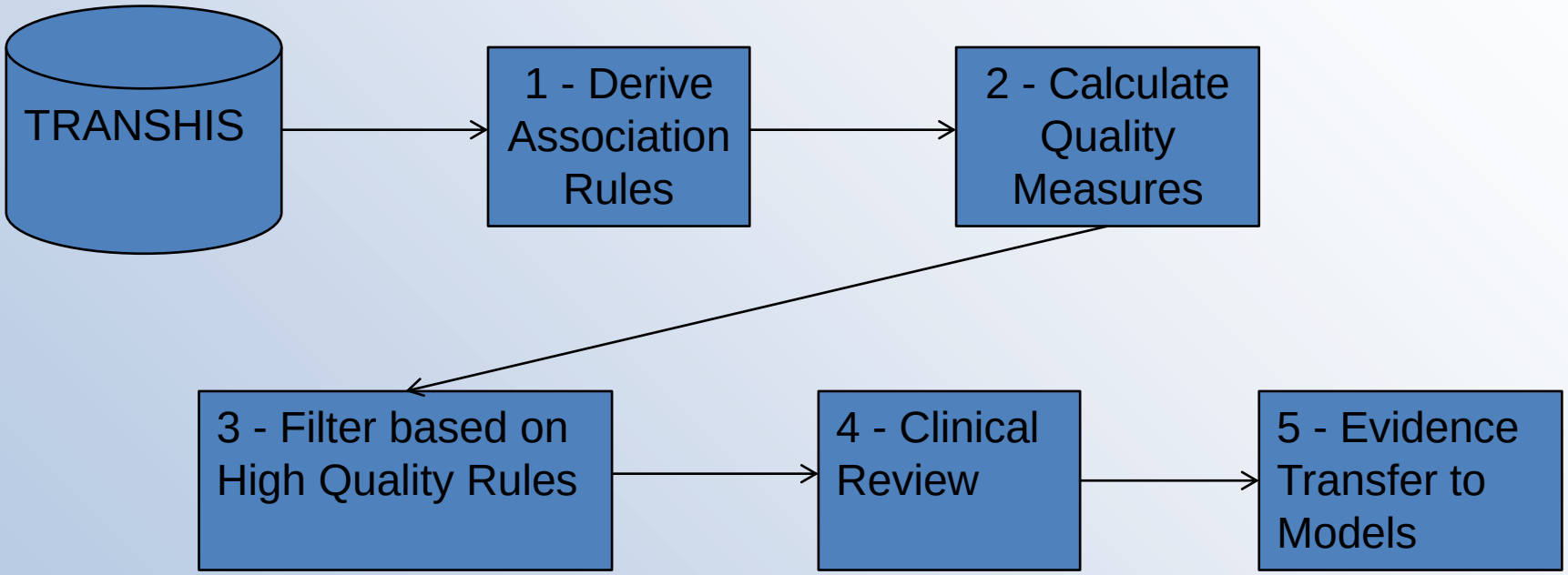
Why data mine?

• Large repositories of potentially useful coded Primary Care data exist such as the TRANSHIS project

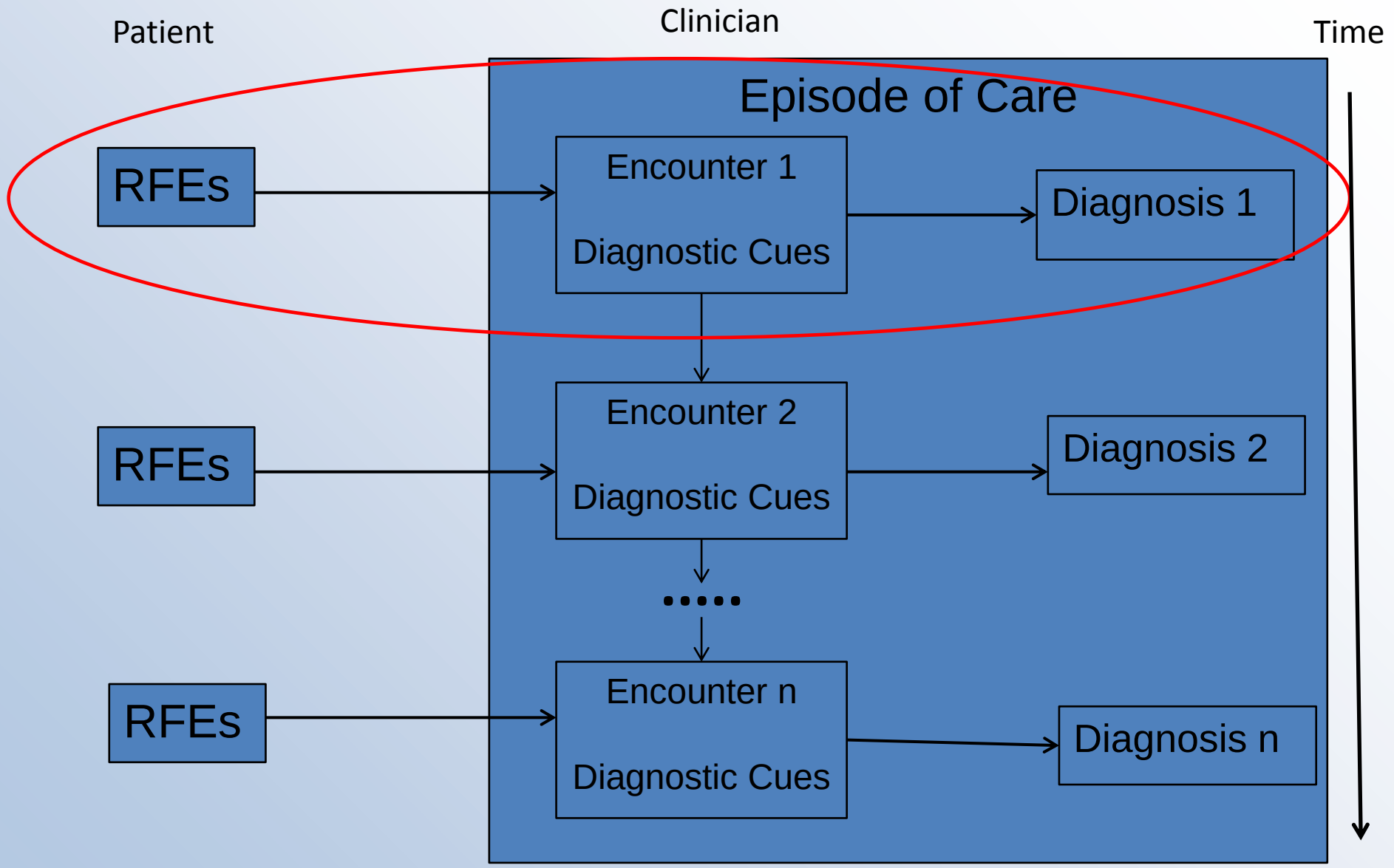
• Can be used to derive empirical evidence, sensitive to different populations, to support the goal of practicing "evidence based medicine"

• In the context of TRANSFoRm, can provide the data necessary to populate actionable models of evidence from dynamic sources, rather than static literature based sources

Data mining steps

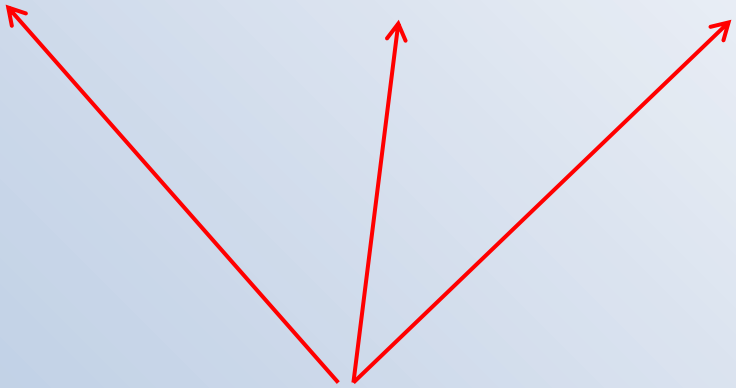


How to data mine TRANSHIS?



Step 1 - Association Rules Structure

- *RFEs, Diagnostic Cues, Demographic Features --------> Diagnosis*



Antecedent Variables --------- > Consequent Variable

e.g. Abdominal Pain, Dysuria, Fever, Female → Urinary Tract Infection

ICPC2 Coded = D06, U01, A03, F -> U70

– Apriori Algorithm implemented using tool called KNIME

Step 2 - General Quality Measures

- Itemsets characterization
  - Support – no of cases containing rule antecedents
- Rules characterization
  - Lift – how much more likely antecedents and consequent occur together than if statistically independent
  - Confidence – probability of consequent occurring given the antecedent

ANTECENDENTS	=>	CONSECUTIVE	AVG SCORE	SUPPOF▼	CONFIDENCE (%)	LIFT	SPECIFITY	SENSITIVITY	LR+	LR-	ODDS	PROVENANCE
D06, U01	=>	U71	0.0	72	58,50	25,50	1,00	0,01	60,10	0,99	60,10	Netherlans
D06, U01, F	=>	U71	0.0	65	62,50	27,23	1,00	0,01	70,95	0,99	70,95	Netherlans
D06, U02	=>	U71	0.0	41	37,60	16,39	1,00	0,01	25,67	0,99	25,67	Netherlans

Step 2 -Bayesian Quality Measures

- Consequent (disease) interest
  - Prior probability – prevalence of disease
- Variables characterization
  - Likelihood ratios (+/-)

ANTECENDENTS	=>	CONSECUTIVE	AVG SCORE	SUPPOF▼	CONFIDENCE (%)	LIFT	SPECIFITY	SENSITIVITY	LR+	LR-	ODDS	PROVENANCE
D06, U01	=>	U71	0.0	72	58,50	25,50	1,00	0,01	60,10	0,99	60,10	Netherlans
D06, U01, F	=>	U71	0.0	65	62,50	27,23	1,00	0,01	70,95	0,99	70,95	Netherlans
D06, U02	=>	U71	0.0	41	37,60	16,39	1,00	0,01	25,67	0,99	25,67	Netherlans

Step 3, 4 -Web rule review tool

- Goal: Analyse data mining results from TRANSHIS to identify clinically significant rules

Features:

- Full rule view
- Sorting rules
- Multivariate filtering by quality measures
- Annotation of interesting findings
- Transfer of rules to evidence ontology service

Antecedents - RfEs

D04
D05
D07
D08
D09
D10

>>
<<

D06

Antecedents - Anams

A01
A02
A03
A04
A05
A06

>>
<<

Sex

Age group

0-4
5-9
10-14
15-19
20-24

Consecutive

U26
U27
U28
U29
U70
U71

Support	Confidence	Lift	Specifity	Sensitivity	LR+	LR-	Odds	Score	Provenance	Scenario
									Netherlans	a_130613

Hide filters | Filter | Clear

Expand table

Number of rules: 43

ANTECENDENTS	=>	CONSECUTIVE	AVG SCORE	SUPPOF	CONFIDENCE (%)	LIFT	SPECIFITY	SENSITIVITY	LR+	LR-	ODDS	PROVENANCE	SCENARIO	ADD REMARK
D06, U01	=>	U71	0.0	72	58,50	25,50	1,00	0,01	60,10	0,99	60,10	Netherlans	a_130613	add rema
D06, U01, F	=>	U71	0.0	65	62,50	27,23	1,00	0,01	70,95	0,99	70,95	Netherlans	a_130613	add rema
D06, U02	=>	U71	0.0	41	37,60	16,39	1,00	0,01	25,67	0,99	25,67	Netherlans	a_130613	add rema

RFE

ICPC2	DESC
U01	Dysuria/painful urination
D06	Abdominal pain localized, other

EPISODE DIAGNOSE

ICPC2

U71

DESCRIPTION

Cystitis/urinary infection, other

TransHIS analysis example

Rules **x => U71**, (Cystitis/urinary infection, other)

- U01 – Dysuria
- U02 – Urinary Frequency
- U06 – Haematuria
- A03 – Fever
- D06 – Abdominal pain localised

- Important cues compare favourably with literature – e.g. JAMA reviews

Rule transfer to Evidence Model

Antecedents - RfEs: A01, A02, A03, A04, A05, A06 (>> <<)

Antecedents - Anams: A01, A02, A03, A04, A05, A06 (>> <<)

Sex

Age group: 0-4, 5-9, 10-14, 15-19, 20-24

Consecutive: U26, U27, U28, U29, U70, U71

Support	Confidence	Lift	Specifity	Sensitivity	LR+	LR-	Odds	Score	Provenance	Scenario
>									Malta	a_130325
10										

Hide filters | Filter | Clear

Collapse table

Number of rules: 42

NSECUTIVE	AVG SCORE	SUPPORT	CONFIDENCE (%)	LIFT ▼	SPECIFITY	SENSITIVITY	LR+	LR-	ODDS	PROVENANCE	SCENARIO	ADD REMARK	TO SEND
1	0.0	12	100,00	78,07	1,00	0,04	1797693134	0,96	179769	Malta	a_130325	add remark	deploy
1	0.0	24	100,00	78,07	1,00	0,07	1797693134	0,93	179769	Malta	a_130325	add remark	deploy
1	0.0	11	100,00	78,07	1,00	0,03	1797693134	0,97	179769	Malta	a_130325	add remark	deploy
1	0.0	24	100,00	78,07	1,00	0,07	1797693134	0,93	179769	Malta	a_130325	add remark	deploy
1	0.0	37	94,90	74,06	1,00	0,11	1425,75	0,89	1425,75	Malta	a_130325	add remark	deploy
1	0.0	17	94,40	73,73	1,00	0,05	1310,15	0,95	1310,15	Malta	a_130325	add remark	deploy
1	0.0	15	93,70	73,19	1,00	0,04	1156,01	0,96	1156,01	Malta	a_130325	add remark	deploy
1	0.0	14	93,30	72,86	1,00	0,04	1078,94	0,96	1078,94	Malta	a_130325	add remark	deploy
1	0.0	13	92,90	72,49	1,00	0,04	1001,88	0,96	1001,88	Malta	a_130325	add remark	deploy
1	0.0	66	91,70	71,56	1,00	0,19	847,74	0,81	847,74	Malta	a_130325	add remark	deploy
1	0.0	15	88,20	68,88	1,00	0,04	578,01	0,96	578,01	Malta	a_130325	add remark	deploy
1	0.0	30	88,20	68,88	1,00	0,09	578,01	0,91	578,01	Malta	a_130325	add remark	deploy

DEPLOY FILTERED

Rule transfer to Ontology

RULE	REMOVE
U02,U01,M => U71	remove
U02,F,U01 => U71	remove
25-29,U01 => U71	remove
U02,F,D06 => U71	remove
D06 => U71	remove
25-29,U02,F => U71	remove
U02 => U71	remove
F,20-24,U01 => U71	remove
25-29,F,D06 => U71	remove
U02,U01 => U71	remove
U01,D06 => U71	remove
25-29,F,U01 => U71	remove
U02 => U71	remove
U02,20-24,U01 => U71	remove
F,U01,50-54 => U71	remove

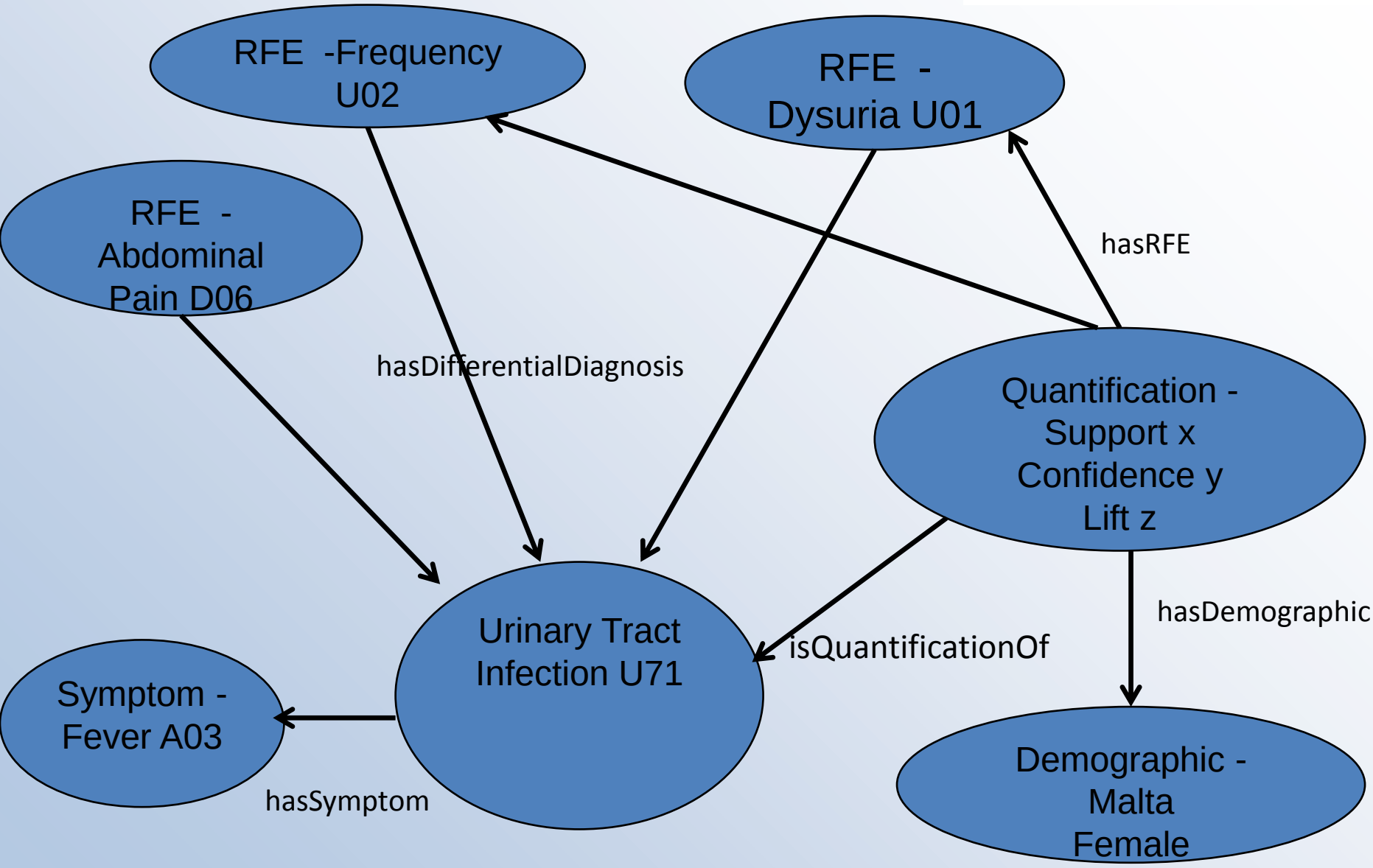
Path to save XML file

Send

Clear

Evidence Representation

U02,F,U01 => U71
A03,D06 => U71



Conclusions

- Feasible to identify clinically meaningful evidence based on coded primary care repositories of data

- Primary Care context is crucial - dependent on how common the condition occurs in primary care e.g. UTI vs Ectopic Pregnancy

- May become more feasible for uncommon cases when data from more countries are aggregated together to give larger volumes of data for uncommon cases

- The TRANSFoRm evidence models can be used to represent and query this data for decision support

Derek Corrigan:

derekcorrigan@rcsi.ie

www.hrbcentreprimarycare.ie

www.transformproject.eu