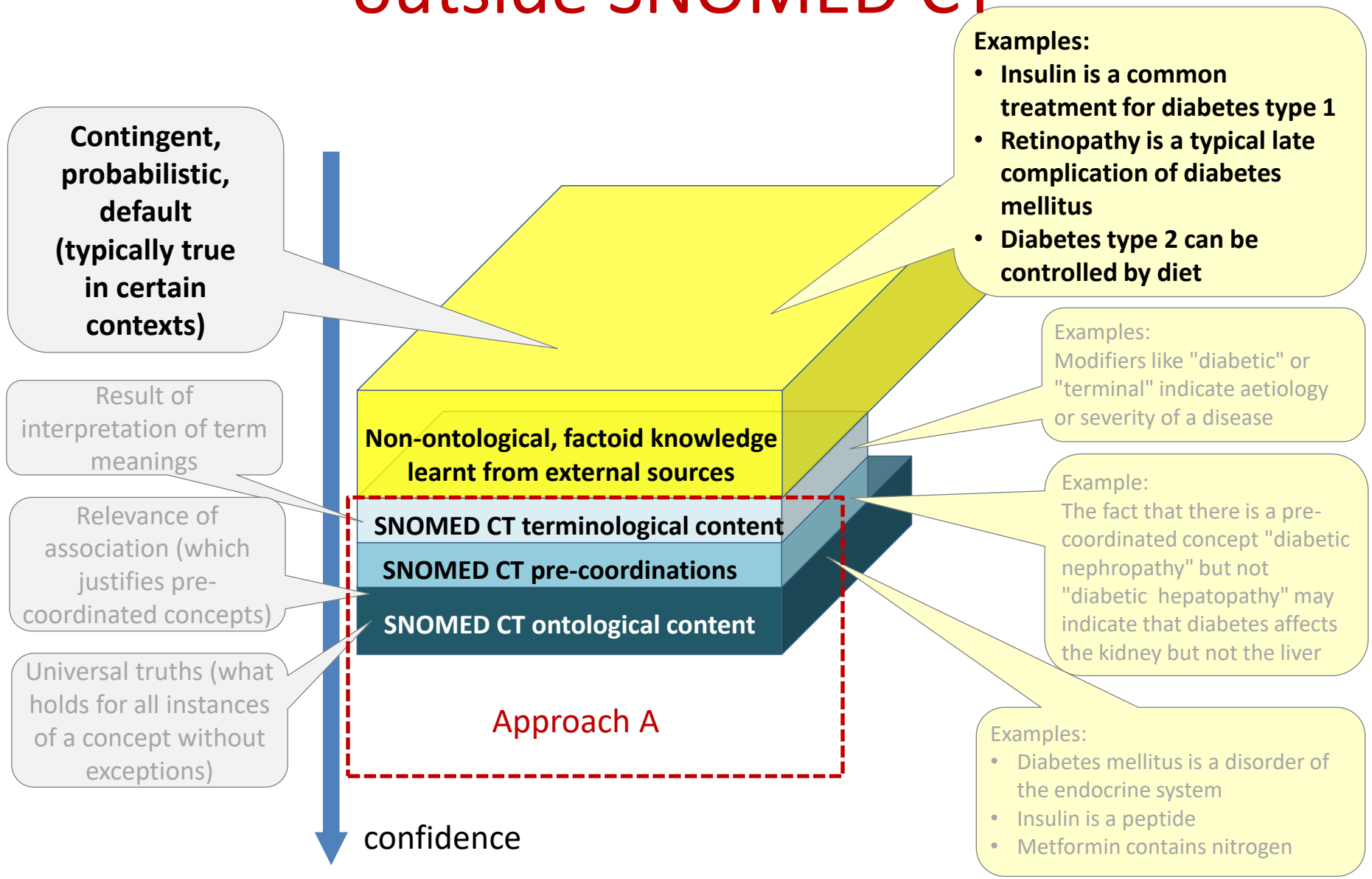
Exploitation of Structured Knowledge Sources for Question Answering: Future Aspects

Stefan Schulz
Markus Kreuzthaler
Ulrich Andersen

Find non-ontological predications outside SNOMED CT



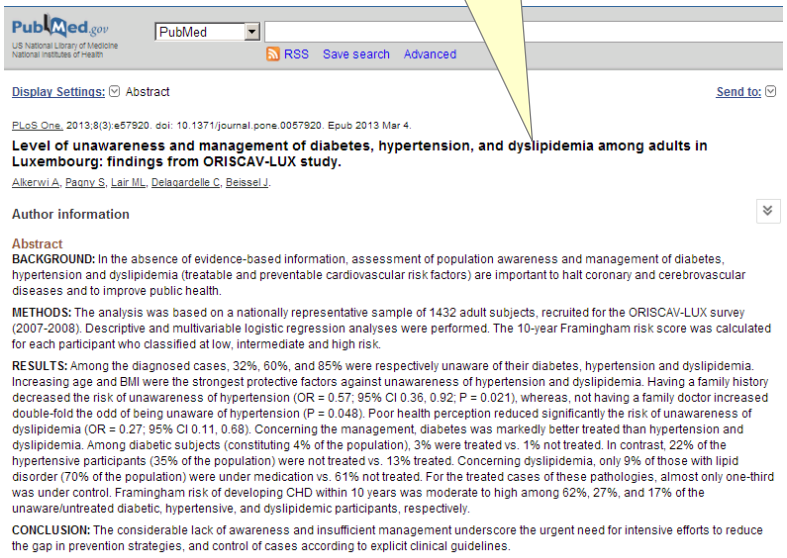
Possible predicates between SNOMED semantic types

	Disease	Finding	Substance	Organism	Body Part	Procedure	Device
Disease	complicates	has sign or symptom	is caused by, is treated by, is prevented by, has metabolite	is caused by, exhibits	is located in	is treated by, is caused by	is caused by
Finding	sign or symptom of	accompanied by	is caused by	is caused by	is located in	is targeted by	is caused by
Substance	causes, treats, prevents, is metabolite of	causes, treats, prevents	interacts	is affected by	targets	is used by	
Organism	causes, is observed in	causes	is sensitive to	interacts	targets		
Body part	is a location of	is a location of	is targeted by	is targeted by			is targeted by
Procedure	treats, causes	treats, causes	uses				uses
Device	causes	causes			targets	is used in	is used with

- most of them are non-ontological and therefore not asserted in SNOMED CT
- Knowledge source to be explored: **UMLS co-occurrence matrix**

Example MEDLINE MeSH annotations

Scientific paper



PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed

RSS Save search Advanced

Display Settings: Abstract Send to:

PLoS One. 2013;8(3):e57920. doi: 10.1371/journal.pone.0057920. Epub 2013 Mar 4.

Level of unawareness and management of diabetes, hypertension, and dyslipidemia among adults in Luxembourg: findings from ORISCAV-LUX study.

Alkerwi A, Pagny S, Lair ML, Delagardelle C, Beissel J.

Author information

Abstract

BACKGROUND: In the absence of evidence-based information, assessment of population awareness and management of diabetes, hypertension and dyslipidemia (treatable and preventable cardiovascular risk factors) are important to halt coronary and cerebrovascular diseases and to improve public health.

METHODS: The analysis was based on a nationally representative sample of 1432 adult subjects, recruited for the ORISCAV-LUX survey (2007-2008). Descriptive and multivariable logistic regression analyses were performed. The 10-year Framingham risk score was calculated for each participant who classified at low, intermediate and high risk.

RESULTS: Among the diagnosed cases, 32%, 60%, and 85% were respectively unaware of their diabetes, hypertension and dyslipidemia. Increasing age and BMI were the strongest protective factors against unawareness of hypertension and dyslipidemia. Having a family history decreased the risk of unawareness of hypertension (OR = 0.57; 95% CI 0.36, 0.92; P = 0.021), whereas, not having a family doctor increased double-fold the odd of being unaware of hypertension (P = 0.048). Poor health perception reduced significantly the risk of unawareness of dyslipidemia (OR = 0.27; 95% CI 0.11, 0.68). Concerning the management, diabetes was markedly better treated than hypertension and dyslipidemia. Among diabetic subjects (constituting 4% of the population), 3% were treated vs. 1% not treated. In contrast, 22% of the hypertensive participants (35% of the population) were not treated vs. 13% treated. Concerning dyslipidemia, only 9% of those with lipid disorder (70% of the population) were under medication vs. 61% not treated. For the treated cases of these pathologies, almost only one-third was under control. Framingham risk of developing CHD within 10 years was moderate to high among 62%, 27%, and 17% of the unaware/untreated diabetic, hypertensive, and dyslipidemic participants, respectively.

CONCLUSION: The considerable lack of awareness and insufficient management underscore the urgent need for intensive efforts to reduce the gap in prevention strategies, and control of cases according to explicit clinical guidelines.

MeSH Main headings

MH - Adolescent
MH - Adult
MH - Aged
MH - Coronary Disease/epidemiology
MH - Diabetes Mellitus/epidemiology/therapy
MH - Dyslipidemias/epidemiology/therapy
MH - Female
MH - Health Knowledge, Attitudes, Practice
MH - Humans
MH - Hypertension/epidemiology/therapy
MH - Luxembourg/epidemiology
MH - Male
MH - Middle Aged
MH - Multivariate Analysis
MH - Prevalence
MH - Risk Factors
MH - Young Adult

MeSH subheadings

MEDLINE bibliographic records (> 20,000,000) are manually annotated using MeSH descriptors

On MEDLINE concept / concept co-occurrences

- The UMLS provides a co-occurrence matrix

Concept 1

Concept 2

records in which C1 and C2 co-occur

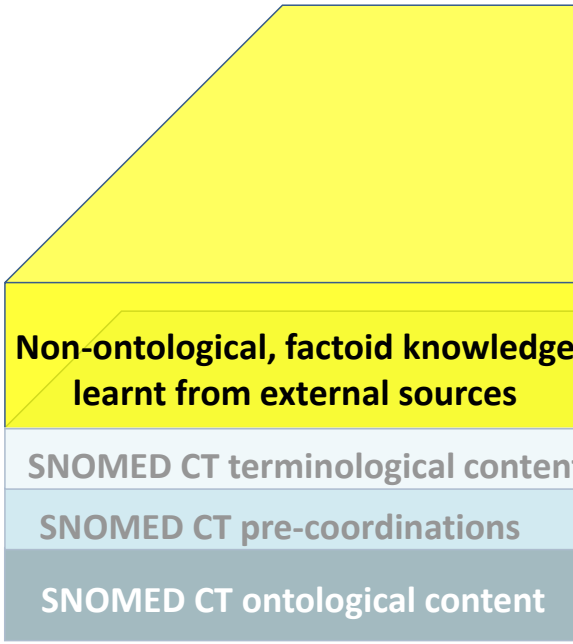
```
C0026683|C0001883|5|CO=4,DI=2,PA=2,SU=2,RA=1
C0026683|C0001948|1|ET=1,SU=1
C0026683|C0002475|1|ET=1,PA=1,TH=1
C0026683|C0003392|1|DT=1
C0026683|C0003466|2|SU=2,DI=1,EP=1,ET=1,PA=1
C0026683|C0003611|5|SU=4,DI=2,PA=2,CO=1,RA=1
C0026683|C0003611|10|SU=8,DI=5,PA=4,CO=3,RA=2,US=2,ET=1
C0026683|C0003614|19|SU=11,PA=10,DI=9,ET=6,CO=5,RA=2,US=2
C0026683|C0003614|21|SU=13,DI=9,CO=8,PA=7,RA=6,ET=5,US=2
C0026683|C0003615|4|SU=4,US=3,PA=2,RA=2,CO=1,DI=1
C0026683|C0003615|5|SU=3,CO=2,PA=2,RA=2,DI=1,US=1
C0026683|C0003617|41|DI=23,SU=23,PA=17,US=12,CO=10,ET=5,RA=4,EP=1
```

MeSH subheadings, which refine the meaning of C1.
E.g.
CO = complicates PA = pathology
DI = diagnoses SU = surgery

Induction of SPO triples by MeSH subheading analysis

- Principle: define filtering conditions for each predicate type
  - Semantic types of concepts (mapped to SNOMED CT)
  - Co-occurrence values
  - Subheading distribution

  } checked against thresholds
- Example: Criteria for: <C1; is treated by; C2>:
  - C1 is of the SNOMED type *Disease or Finding*
  - C2 is of one of the types *Substance*, *Product*, *Device*, *Procedure*
  - C1 / C2 co-occurrence above threshold
    - log-likelihood > 6.63, corresponds to p<0.01
  - thresholds of subheading rates
    - DT (drug therapy) > 50% or
    - DH (diet therapy) > 50% or
    - TH (therapy) > 50%
- Implemented: Java + Lucene



"What can be done to prevent hyperglycemia?"

```
(Hyperglycemia (disorder),Prevents (attribute),prevents,*ANY)

Question: "Hvad er vigtigt For at forebygge alvorlige tilfælde af hyperglykæmi?","Tree": "(ROOT (subj (PRON (Hvad hvad)))
(VERB (er være)) (pred (ADJ (vigtigt vigtig)) (pobj (ADP (For for)) (nobj (X (at at)) (vobj (VERB (forebygge forebygge))
(dobj (mod (ADJ (alvorlige alvorlig))) (NOUN (tilfælde tilfælde)) (pobj (ADP (af af)) (nobj (NOUN (hyperglykæmi
hyperglykæmi)))))))) (pnct (X (? ?))))","hierarchy1": "Clinical Finding","concept1ID": "80394007","term1":
"Hyperglykæmi","attribute": "116699007|prevents","hierarchy2": "*","concept2ID": "*hvad","term2":
"hvad","ESICT_EXPRESSION": "(80394007,116699007|prevents,*hvad)","trace": "R14","jasvar": "","nejsvar":
"","forklarendesvar": "Vigtigt for at forebygge alvorlige tilfælde af hyperglykæmi er følgende"

Trigger: ASSOCIATED WITH

Generated Lucene Query:
IS PREVENTED BY
sidOne:80394007 AND semTypeOne:(disorder OR finding) AND PC:[00000050 TO 00000100]

Command Line:
java -jar esict.jar tmp c:\DataESICT\luceneIndexCoocDataLogLikeFullSubHeading
sidOne:80394007_AND_semTypeOne:(disorder_OR_finding)_AND_PC:[00000050_TO_00000100]  20

Answers:
Hyperglycemia (disorder) IS PREVENTED BY Perioperative care (regime/therapy) log-like: 67,29
Hyperglycemia (disorder) IS PREVENTED BY Artificial pancreas, device (physical object) log-like: 47,22
Hyperglycemia (disorder) IS PREVENTED BY Dietary fiber (substance) log-like: 24,34
Hyperglycemia (disorder) IS PREVENTED BY Human insulin (substance) log-like: 22,52
Hyperglycemia (disorder) IS PREVENTED BY Human insulin product (product) log-like: 22,52
Hyperglycemia (disorder) IS PREVENTED BY Nutritional support (regime/therapy) log-like: 18,64
Hyperglycemia (disorder) IS PREVENTED BY 1 Deoxynojirimycin (substance) log-like: 15,47
Hyperglycemia (disorder) IS PREVENTED BY 1 Deoxynojirimycin (product) log-like: 15,47
Hyperglycemia (disorder) IS PREVENTED BY Saccharomyces cerevisiae (organism) log-like: 14,14
Hyperglycemia (disorder) IS PREVENTED BY Glyburide (product) log-like: 13,33
Hyperglycemia (disorder) IS PREVENTED BY Glyburide (substance) log-like: 13,33
Hyperglycemia (disorder) IS PREVENTED BY Intraoperative care (regime/therapy) log-like: 12,96
Hyperglycemia (disorder) IS PREVENTED BY Gliclazide (substance) log-like: 7,79
```

"How can diabetes mellitus be treated?"

```
1)
(73211009,116700008|treats,243120004)
(Diabetes mellitus (disorder),Treats (attribute)|treats,Regimes and therapies (regime/therapy))

Question: "Hvad bygger Diabetes behandling på?","Tree": "(ROOT (VERB (bygger bygge)) (dobj (NOUN (Diabetes diabetes))
(possd (NOUN (behandling behandling)) (pobj (nobj (PRON (Hvad hvad))) (ADP (på på))))) (pnct (X (? ?))))","hierarchy1":
"Clinical Finding","concept1ID": "73211009","term1": "diabetes mellitus ","attribute": "116700008|treats","hierarchy2":
"Procedure","concept2ID": "243120004","term2": "regimer og behandlinger","ESICT_EXPRESSION":
"(73211009,116700008|treats,243120004)","trace": "R14","jasvar": "","nejsvar": "","forklarendesvar": "Diabetes
behandling bygger på følgende"

Trigger: ASSOCIATED WITH
Generated Lucene Query:
IS TREATED BY
sidOne:73211009 AND semTypeOne:(disorder OR finding) AND DT:[00000050 TO 00000100] OR TH:[00000050 TO 00000100] OR
DH:[00000050 TO 00000100]
Command line:
java -jar esict.jar tmp c:\DataESICT\luceneIndexCoocDataLogLikeFullSubHeading
sidOne:73211009_AND_semTypeOne:(disorder_OR_finding)_AND_semTypeTwo:(substance_OR_product_OR_device_OR_procedure)_AND_(D
T:[00000050_TO_00000100]_OR_TH:[00000050_TO_00000100]_OR_DH:[00000050_TO_00000100]) 20

Answers:
Diabetes mellitus (disorder) IS TREATED BY Hypoglycemic agent (substance) log-like: 3.387,74
Diabetes mellitus (disorder) IS TREATED BY Hypoglycemic agent (product) log-like: 3.387,74
Diabetes mellitus (disorder) IS TREATED BY Regular insulin (substance) log-like: 2.420,36
Diabetes mellitus (disorder) IS TREATED BY Insulin (substance) log-like: 2.420,36
Diabetes mellitus (disorder) IS TREATED BY Insulin product (product) log-like: 2.420,36
Diabetes mellitus (disorder) IS TREATED BY Antineoplastic agent (substance) log-like: 164,00
Diabetes mellitus (disorder) IS TREATED BY Antineoplastic agent (product) log-like: 164,00
Diabetes mellitus (disorder) IS TREATED BY Thiazolidinedione (substance) log-like: 157,26
Diabetes mellitus (disorder) IS TREATED BY Thiazolidinedione (product) log-like: 157,26
Diabetes mellitus (disorder) IS TREATED BY Metformin (substance) log-like: 147,59
Diabetes mellitus (disorder) IS TREATED BY Metformin (product) log-like: 147,59
Diabetes mellitus (disorder) IS TREATED BY Glyburide (substance) log-like: 145,51
Diabetes mellitus (disorder) IS TREATED BY Glyburide (product) log-like: 145,51
Diabetes mellitus (disorder) IS TREATED BY Sulfonylurea and its derivatives (substance) log-like: 132,04
Diabetes mellitus (disorder) IS TREATED BY Sulfonylurea (substance) log-like: 132,04
Diabetes mellitus (disorder) IS TREATED BY Sulfonylurea (product) log-like: 132,04
Diabetes mellitus (disorder) IS TREATED BY Subcutaneous injection (procedure) log-like: 79,87
Diabetes mellitus (disorder) IS TREATED BY Long-acting insulin (substance) log-like: 75,22
Diabetes mellitus (disorder) IS TREATED BY Long acting insulin (product) log-like: 75,22
```

Outlook

- Approach included into ESICT interface soon
- Known limitations
  - UMLS COOC table lacks important information from MEDLINE (document type, non-human, chemicals)
  - Low granularity of MeSH compared to SNOMED CT
  - Cooccurrences not aggregated in the hierarchy
  - No distinction between hypotheses studied and scientific evidence
- Possible future work:
  - Using MEDLINE source data
  - Using text-mined content from abstracts