Vaccine Semantics

Automatic methods for recognizing, representing, and reasoning about vaccine-related information

Benedikt Becker
January 8, 2019

In the news:

Rare side events after authorization?

CBS NEWS   NEWS   SHOWS   LIVE

GBS a Side Effect of H1N1 Vaccine?

NOV 20, 2009 / CBS

Health officials say they are carefully monitoring the H1N1 vaccine program for any dangerous side effects, including a very rare syndrome known as Guillain-Barre syndrome, also known as GBS.

However, **CBS News** medical correspondent **Dr. Jennifer Ashton** reported GBS has been cropping up this year, possibly in connection with the H1N1 vaccine. She shared the story of one young man who developed GBS, and why his parents think it may be related to the H1N1 vaccine.

Jordan McFarland, an athletic 14-year-old, was weeks ago playing tennis and basketball. Now he needs a walker to move from room to room.

McFarland told **CBS News**, "It's an aching, but it's also a pain that I can't describe."

In the news:

Rare side events after authorization?

CBS NEWS NEWS SHOWS

The New York Times

Prepare for a Vaccine Controversy

By ARTHUR ALLEN AUG. 1, 2009

Washington

A FEW years ago public health officials set up a time share in Pennsylvania hens. Under contracts signed with several farmers, the hens continued to lay for their regular customers until the moment this past spring when the federal government requisitioned their eggs to grow flu vaccine.

Strategic hen reserves are part of a success story: the government's readiness for the current H1N1 flu pandemic. Public health officials had already stockpiled millions of doses of antiviral drugs, created diagnostic kits that detected the virus as soon as it appeared in California in April and enrolled five companies to make vaccine. By mid-October we may have as many as 80 million doses ready for a mass immunization program.

In the news:

Rare side events after authorization?

NEWS SHOWS

The New York Times

BMJ. 2011; 343: d3908.

Guillain-Barré syndrome and adjuvanted pandemic influenza A (H1N1) 2009 vaccine: multinational case-control study in Europe

Abstract

Objective To assess the association between pandemic influenza A (H1N1) 2009 vaccine and Guillain-Barré syndrome.

Design Case-control study.

Setting Five European countries.

Participants 104 patients with Guillain-Barré syndrome and its variant Miller-Fisher syndrome matched to one or more controls. Case status was classified according to the Brighton Collaboration definition. Controls were matched to cases on age, sex, index date, and country.

Main outcome measures Relative risk estimate for Guillain-Barré syndrome after pandemic influenza vaccine.

Results Case recruitment and vaccine coverage varied considerably between countries; the most common vaccines used were adjuvanted (Pandemrix and Focetria). The unadjusted pooled risk estimate for all countries was 2.8 (95% confidence interval 1.3 to 6.0). After adjustment for influenza-like illness/upper respiratory tract infection and seasonal influenza vaccination, receipt of pandemic influenza vaccine was not associated with an increased risk of Guillain-Barré syndrome (adjusted odds ratio 1.0, 0.3 to 2.7). The 95% confidence interval shows that the absolute effect of vaccination could range from one avoided case of Guillain-Barré syndrome up to three excess cases within six weeks after vaccination in one million people.

Conclusions The risk of occurrence of Guillain-Barré syndrome is not increased after pandemic influenza vaccine, although the upper limit does not exclude a potential increase in risk up to 2.7-fold or three excess cases per one million vaccinated people. When assessing the association between pandemic influenza

Vaccines are special

Vaccine success

- among the most effective means for improving population health
- e.g., smallpox, polio, measles

Risk of adverse events

- administered to healthy persons
- requires careful consideration of benefits and risks

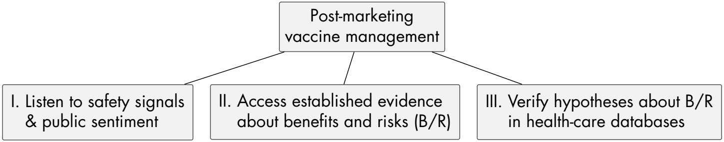
Tests before marketing

- selective populations
- limited follow-up
- time pressure for seasonal vaccines

Post-marketing benefits and risk

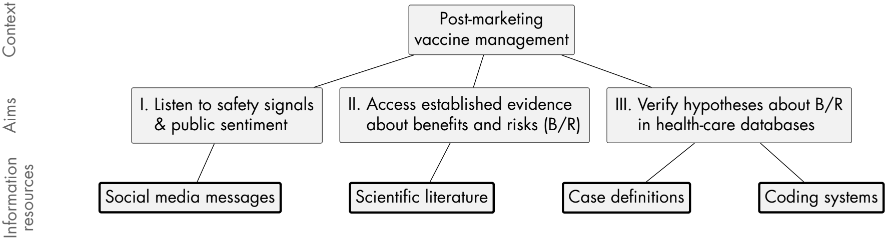
- rare and long-term adverse events
- changes in effectiveness and burden of disease
- possibly strong dynamics in public sentiment

Thesis overview



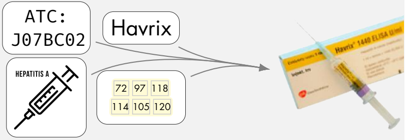
- speed of information retrieval is fundamental
- thesis objective: acceleration by automation

Thesis overview



Representational heterogeneity

Symbol $\nleftrightarrow$ Concept (or Relation)



Resolution steps: Vaccine Semantics

1. recognize the symbols that carry relevant information
2. represent the information independently from its symbols
3. reason about the information using domain knowledge

Thesis overview

Context

- Post-marketing vaccine management

Aims

- I. Listen to safety signals & public sentiment
- II. Access established evidence about benefits and risks (B/R)
- III. Verify hypotheses about B/R in health-care databases

Information resources

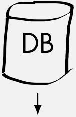
- Social media messages (I)
- Scientific literature (II)
- Case definitions (III)
- Coding systems (III)

Approaches

A. Task-specific rules

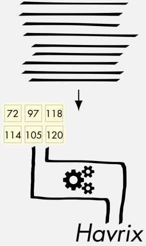
1. `if` 72 97 118 114 105 120 `then` *Havrix*
2. ...

B. Formalized domain knowledge

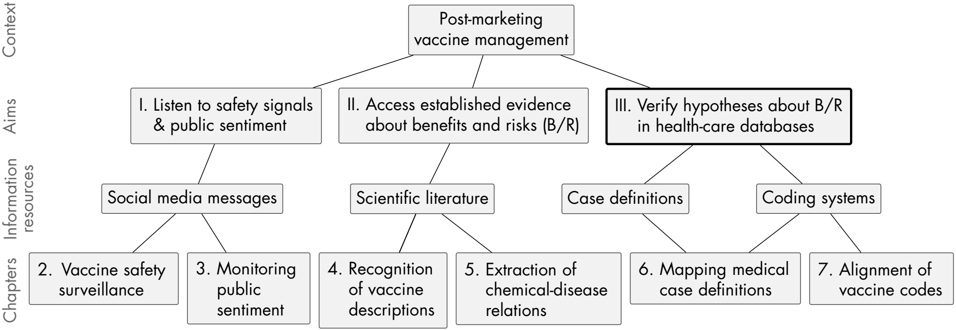


Subject	Predicate	Object
Havrix	*represented-by*	72 97 118 114 105 120
...		

C. Machine learning



Thesis overview



Information extraction toolbox

- natural language processing
- machine learning
- ontology design
- automatic reasoning

Part III: Verification of hypotheses about vaccines in health-care databases

- hypotheses about the benefits and risks of vaccines tested mostly by observational studies in health-care databases
  - primary care, hospitalizations, reimbursement, ...
  - based on identification of vaccinations and medical events (vaccine-preventable disease/adverse events)
- information stored using medical coding system, e.g. ICD-10:

Code	Descriptor
J13	Pneumonia due to Streptococcus pneumoniae
J18	Pneumonia, unspecified organism
J18.0	Bronchopneumonia, unspecified
...	

- increase study scale by combining data from multiple health-care databases

Representation of medical information in Europe

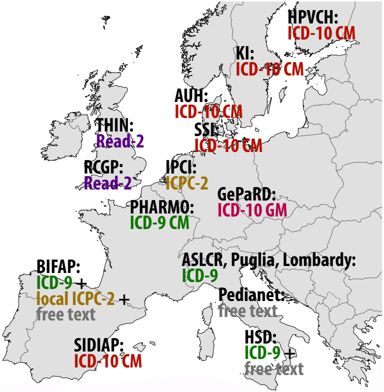
Medical events

- various standardized coding systems
- e.g. codes for *Pneumonia*:

ICD-10 CM: J17, J13, J12, J14, J15, J16, J18, ...
ICD-9 CM: 486, 480, 482.3, 482.9, 487.0, 483, 482.2, 481, 485
ICPC-2: R81
Read-2: H25.., H222., H22z., H26.., H22yz, H23.., H2700, H20.., H223.

Vaccines

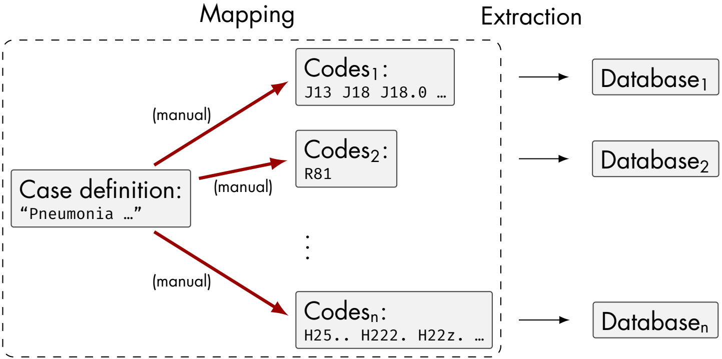
- custom, database-specific coding systems
  - descriptors often in national languages



ADVANCE project (2013-2019)

- Accelerated development of vaccine benefit-risk collaboration in Europe

Naive approach: manual creation of code sets

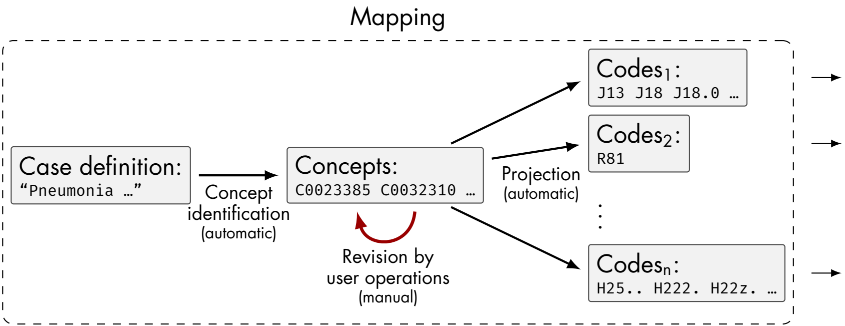


Drawbacks for collaborative studies

- creation of individual extraction queries requires extensive manual work
- no reinforcement of consistency between extraction queries

Ch. 6: CodeMapper: Semi-automatic mapping of case definitions

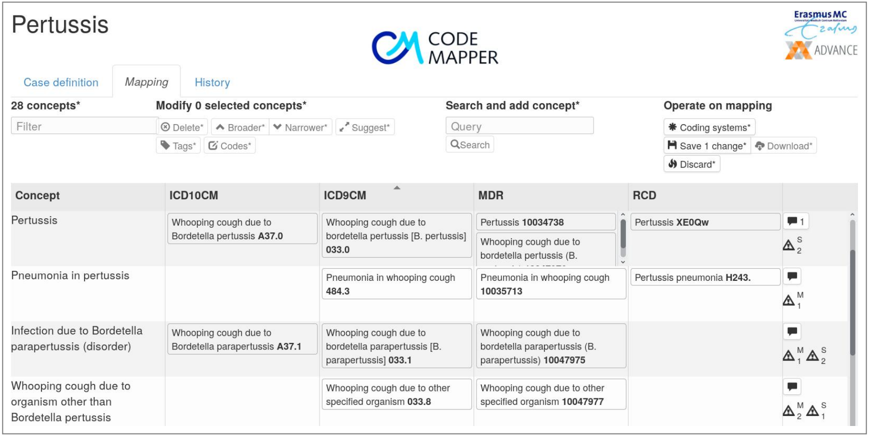
Objective: map textual case definition to database codes with minimal manual effort



- automation and user operations using the Unified Medical Coding System (UMLS), based on Avillach [2013]
  - ensures consistency of code sets
  - mapping process independent of targeted coding systems
- evaluation showed effectiveness of user operations and necessity of human revision

Ch. 6: CodeMapper: Semi-automatic mapping of case definitions

- open-source web application at `https://euadr.erasmusmc.nl/CodeMapper`
- tracking of mapping process
- applied in projects ADVANCE and EMIF, and in industry



Ch. 7: Alignment of vaccine codes using the VaccO ontology of vaccine descriptions

Application in vaccine studies

- ▷ reference vaccine coding system to specify vaccine or vaccine group
- ▷ alignment with database coding systems to identify vaccinations

Reference coding system

Code	Descriptor
...	
J07BC	Hepatitis vaccines
J07BC01	Hepatitis B, purified antigen
J07BC02	Hepatitis A, inactivated, whole virus
...	

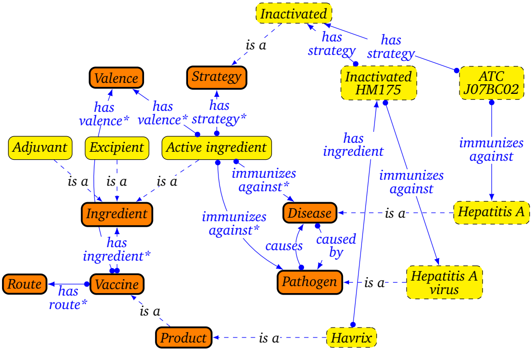
Database coding system

Code	Descriptor
...	
VHM175	Vaccine using HM175
...	

Code alignment

- ▷ identify for each reference code the closest corresponding database code(s)
- ▷ based on ontology alignment [Euzenat & Shvaiko, 2013]

VaccO ontology of vaccine descriptions



- formalization of properties used in vaccine descriptions
  - allows formalization of vaccine descriptions
  - ontology reasoner for inference using domain knowledge

Havrix ≡ *Product that has-ingredient HM175* ⇒ *Havrix is-a Vaccine that imm.-against Hepatitis-A*

- 1,019 classes with 2,962 terms each in up to five languages

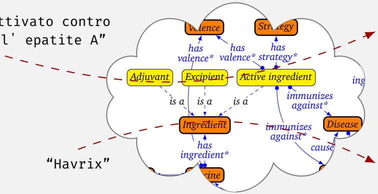
Ontology-based code alignment

1. compute for all database and reference codes a flat representation of available information

Code descriptors *VaccO ontology* *Inferred properties*

“vaccino inattivato contro il virus dell’epatite A”

“Havrix”



Property	Values
Imm. target	Hepatatis A, Hepatitis, Virus
Strategy	Inactivated

Property	Values
Imm. target	Hepatatis A, Hepatitis, Virus
Strategy	Inactivated
Ingredient	HM107

2. measure similarity between codes by overlap between inferred properties
3. assign every database code to the reference code with maximal similarity above threshold

Ch. 7: Alignment of vaccine codes using the VaccO ontology

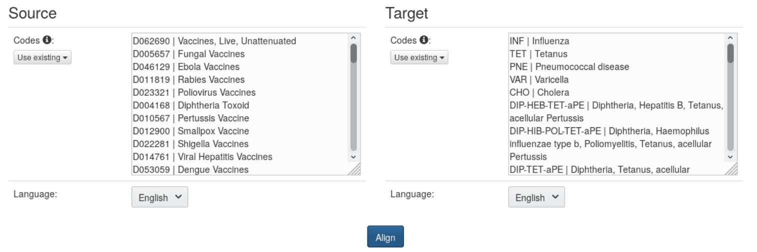
Evaluation and application

Evaluation

- two existing reference sets of manually created vaccine code alignments (ADVANCE Vactype, UMLS/ATC)
- calculate F-score for re-creating alignments
- excellent performance in both reference sets (avg. 91% and 96% F-score)

Application

- open-source web application at `https://euadr.erasmusmc.nl/VaccO/`



Conclusions

Contributions to post-marketing management of vaccines

Part I – public social media

- ▷ negative evidence for monitoring vaccine safety
- ▷ possible use for monitoring public confidence

Part II – scientific literature

- ▷ building blocks for mining vaccine-related information

Part III – observational studies

- ▷ formalized domain knowledge for unifying codes
- ▷ two user applications to help collaborative studies about vaccines

Conclusions

Extraction of heterogeneously represented information about vaccines

Rule-based approaches

- ▷ lack flexibility and scalability for dealing with free text

Machine learning methods

- ▷ largest flexibility for relevant tasks
- ▷ only few training corpora vaccine domain

Formalized domain knowledge

- ▷ costly to create but applicable to many problems in the domain
- ▷ interpretable, correctable, updateable

Vaccine Semantics

Automatic methods for recognizing, representing, and reasoning about vaccine-related information

Benedikt Becker
January 8, 2019