The SeValid project

Giulia Hyeraci
Pharmacist, Pharmacoepidemiology consultant
ARS Toscana

Validation workshop 27 march 2025
ARS, Villa la Quiete

Agenda

- **SeValid Project Overview**
- **SeValid's Contribution: a broader approach to identifying additional cases**
- **Myocarditis Case Study**
- **Lessons Learnt**
- **Considerations for Xabi, Ersilia and Robert**

Agenda

- **SeValid Project Overview**
- **SeValid's Contribution: a broader approach to identifying additional cases**
- **Myocarditis Case Study**
- **Lessons Learnt**
- **Considerations for Xabi, Ersilia and Robert**

SeValid project overview

WHAT IS IT?

Methodological study that aims to estimate the validity, in terms of positive predictive value, specificity, and sensitivity, of algorithms used to identify cases of myocarditis and deep vein thrombosis using hospital discharge forms

WHICH ALGORITHMS?

- Specific algorithm > based on diagnostic codes closely related to the event of interest
- Screening algorithm > [illegible]t hide cases of the event of interest

Aimed at capturing all or nearly all cases

SeValid project overview

OBJECTIVE

To estimate the **sensitivity of "narrow" algorithms** using **"possible" algorithms** as '**screening**' algorithms

Crucial to develop the list of codes to be included into the possible algorithm!!!

IMPLICATIONS

If successful, the SeVa tion of real-world event incidence rates, streng vaccine benefit-risk profiles and improving the reliability of real-world clinical evidence in decision-making

Example

Validation to estimate sensitivity along with positive predictive value: a case study

Giulia Hyeraci[1], Claudia Bartolini[1], Anna Girardi[1], Giorgio Limoncella[1], Giuseppe Roberto[1], Edith Heintjes[2], Hilda de Jong[2], Josine Kuiper[2], Anna Beckmeyer-Borowko[3], Sigrid Behr[3], Miriam Sturkenboom[4], Rosa Gini[1]

1. Agenzia regionale di Sanità della Toscana, Florence, Italy; 2. PHARMO Institute, Utrecht, The Netherlands; 3. Novartis Pharma AG, Basel, Switzerland; 4. University Medical Center Utrecht, Utrecht, The Netherlands

Email: giorgio.limoncella@ars.toscana.it

Background

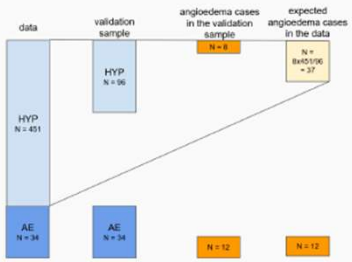
To validate an algorithm retrieving cases of a disease from real-world data sources, positive predictive value (PPV) is often estimated, while sensitivity is rarely assessed. When a sensitive auxiliary algorithm can identify all cases in the population of interest, estimation of sensitivity of the algorithm is possible, using a published analytical formula based on PPV of both algorithms. If this assumption does not hold, the formula estimates the best case scenario for the true sensitivity.

Objectives

To illustrate how PPV of an auxiliary algorithm can be used to estimate sensitivity of an algorithm retrieving cases of angioedema

Methods

Diagnosis of angioedema recorded as ICD9CM 9951 in the emergency room or during hospital admissions between 2016-2019 were extracted from an administrative data source of the Tuscany region in Italy. As an auxiliary algorithm, diagnoses were also extracted compatible with angioedema (hypersensitivity: H), ICD9CM: 9950, 99527, 7088, 7080, 7081, 7089, 37633, 37482, 47825, 4786, 47875, 5088, 7823, 9952. All records of angioedema and a sample of hypersensitivity were validated by two assessors to classify true cases of angioedema.



Positive predictive value was calculated for both algorithms and a best-case scenario estimate of the sensitivity was calculated for the angioedema algorithm.

Main Algorithm: angioedema diagnosis, **A**

Auxiliary Algorithm: hypersensitivity diagnosis, **H**

By exploiting the relationships between the sensitivities (SE) of the main algorithm and the auxiliary algorithm:

$$SE_{A\cap H} = SE_A + SE_H + SE_{A\cap\overline{H}};$$

and the interrelationships between validation indices:

$$\pi = \frac{P \times PPV}{SE}; \quad SE = \frac{P \times PPV}{\pi}; \quad PPV = \frac{SE \times PPV}{\pi};$$

(where: π is the true prevalence and P is the observed prevalence)

we obtain a best-case scenario estimate of the sensitivity of the angioedema algorithm:

$$SE_A \leq \frac{P_A PPV_A}{P_A PPV_A + P_H PPV_H}$$

or:

$$SE_A \leq \frac{N_A PPV_A}{N_A PPV_A + N_H PPV_H} \quad (1)$$

(where: N is the absolute number of individual detected by an algorithm)

However, due to COVID restrictions, assessors could not access medical charts. Diagnoses recorded during the hospitalisation, free text observations during emergency room, and all medicine dispensing recorded up to 30 days following diagnosis were used instead.

Results

Cases identified for angioedema and hypersensitivity were 34 and 451, respectively. Validation resulted in 12 true cases out of 34 for angioedema and 8 in a sample of 96 for hypersensitivity, thus, the estimate of positive predictive values corresponds to:

$$PPV_A = \frac{12}{34} = 0.353$$

$$PPV_H = \frac{8}{96} = 0.083$$

and, using the formula 1:

$$SE_A \leq \frac{34 \times 0.353}{34 \times 0.353 + 451 \times 0.083} = 0.245$$

From a more practical point of view, the difficulty in sensitivity estimation lies in identifying the set of false negatives. If sensitivity of A is defined as:

$$SE_A = \frac{TP_A}{TP_A + FN_A}$$

(where: TP is the number of true positive and FN is the number of false positive) we found a subset of the false negative (SFN) using an auxiliary algorithm. We estimated that in the population there were at least 45.1x8.3% = 37 ($PPV_H \times N_H$) cases of angioedema found by the auxiliary algorithm (hypersentitivity diagnosis) that were false negative for main algorithm (angioedema diagnosis), then the sensitivity of the main algorithm can not be higher than:

$$SE_A = \frac{TP_A}{TP_A + FN_A} \Rightarrow SE_A \leq \frac{TP_A}{TP_A + SFN_A} = \frac{12}{12 + 37} = 0.245$$

Conclusion

Based on this illustrative example, the sensitivity of angioedema in the ARS Tuscany data source was not higher than 24.5%. This methodology can be leveraged in pharmacoepidemiologic studies to provide best case sensitivity of case-finding algorithms. This case study should be considered relevant from a methodological point of view, but not from a clinical perspective, due incomplete case adjudication

SeValid project overview

Myocarditis specific algorithm: main results

- A total of 38 cases of cases were retrieved by the hospital discharge records of Careggi using “narrow” ICD9CM diagnostic codes
  (July 1, 2022, and June 30, 2024)

- A **human** reviewer filled out **38 questionnaires** to extract the level of certainty per each case resulting in **37 CASES** of myocarditis with different level of certainties and **1 NOT ASSESSABLE CASE** due to insufficient information to meet the case definition

ICD9CM	429.0	Myocarditis. unspecified
ICD9CM	422.99	Other acute myocarditis
ICD9CM	422.93	Toxic myocarditis
ICD9CM	422.91	Idiopathic myocarditis
ICD9CM	422.90	Acute myocarditis. unspecified
ICD9CM	422.9	Other and unspecified acute myocarditis
ICD9CM	422.0	Acute myocarditis in diseases classified elsewhere
ICD9CM	422	Acute myocarditis
ICD9CM	398.0	Rheumatic myocarditis
ICD9CM	391.2	Acute rheumatic myocarditis
ICD9CM	130.3	Myocarditis due to toxoplasmosis
ICD9CM	093.82	Syphilitic myocarditis
ICD9CM	074.23	Coxsackie myocarditis
ICD9CM	036.43	Meningococcal myocarditis
ICD9CM	032.82	Diphtheritic myocarditis
ICD9CM	422.92	Septic myocarditis

Agenda

- SeValid Project Overview
- **SeValid's Contribution: a broader approach to identifying additional cases**
- Myocarditis Case Study
- Lessons Learnt
- Considerations for Xabi, Ersilia and Robert

SeValid's contribution

A broader approach to identifying additional cases

Context

Strict coding criteria may not capture all relevant cases

Administrative reasons

Clinical reasons

Lack of sufficient clinical reason at the time of coding

Need for a different perspective

Adopting the coder's viewpoint to better capture relevant cases

SeValid's contribution

A broader approach to identifying additional cases

Main tool: interviews addressed to professionals with experience in the field of recording diagnosis into hospital discharge records

The process step-by step:

1) develop the interview
2) test the interview
3) administer the interview to relevant professional
4) results collection
5) integration of relevant information into possible code list definition

Agenda

- SeValid Project Overview
- SeValid's contribution: a broader approach to identifying additional cases
- **Myocarditis Case Study**
- Lessons Learnt
- Considerations for Xabi, Ersilia and Robert

Myocarditis Case Study

Main sections of the interview:

Part 1: Demographic Information (e.g. medical profession, years of experience….)

Part 2: Practice of coding cases of myocarditis (e.g. experience with myocarditis diagnosis, accuracy and misclassification of codes, completeness of coding system)

Part 3: Other codes (e.g. alternative coding choices and reasons, additional coding relevant codes)

Myocarditis Case Study

Key insights from the interview

Part 1: Demographic Information

Cardiologist profile

- 10 years of experience
- Works in inpatient setting
- Responsible to fill out hospital discharge records
- Specialized in patients hospitalized in the Cardiac Intensive Care Unit

Myocarditis Case Study

Key insights from the interview

Part 2: Practice of coding cases of myocarditis

- **Coding Myocarditis Cases**: rare but usually coded correctly at discharge.
- **Potential Misclassification**: errors stem from incomplete available informatiom, not incorrect code usage.
- **Recording habits**: recording of rare myocarditis types may be missing from standard data.

Myocarditis Case Study

Key insights from the interview

Part 3: Other codes

- **Reasons for choosing alternative codes:** more severe conditions, cost considerations, lack of clinical confirmation at discharge
- **Coding Challenges & Alternatives:**
  Cardiogenic shock, acute heart failure, and intermediate coronary syndrome may conceal myocarditis cases
  Reviewing 500 or 200 cases of unspecified codes would likely reveal myocarditis cases

Myocarditis Case Study

Potential additional codes for detecting cases of myocarditis

Cardiogenic shock	Acute heart failure	Intermediate coronary syndromes
785.51 - cardiogenic shock	**428.21** – Acute systolic heart failure **428.23** – Acute on chronic systolic heart failure **428.31** – Acute diastolic heart failure **428.33** – Acute on chronic diastolic heart failure **428.41** – Acute combined systolic and diastolic heart failure **428.43** – Acute on chronic combined systolic and diastolic heart failure	**411.1** - acute coronary syndrome

Agenda

- SeValid Project Overview
- SeValid's Contribution: a broader approach to identifying additional cases
- Myocarditis Case Study
- **Lessons Learnt**
- Considerations for Xabi, Ersilia and Robert

Lessons learnt

- **Looking beyond the expected:** focusing only on “narrow” cases is not enough
- **Think like a coder:** coders don’t always pick the expected exact code for a diagnosis

- **Power of collaboration:** an interdisciplinary approach helps uncover blind spots and ensures a more accurate study design

Agenda

- SeValid Project Overview
- SeValid's Contribution: a broader approach to identifying additional cases
- Myocarditis Case Study
- Lessons Learnt
- **Considerations for Xabi, Ersilia and Robert**

Considerations for Xabi

- Engage with **those** responsible for **coding key outcomes and variables** to gain usefull insights from the real-world clinical practice

- Develop **screening algorithms** capable of identifying **additional cases** for the event of interest

Considerations for Ersilia and Robert

- How can we adopt the **perspective of coders** to improve the identification of key variables and outcomes in the healthcare administrative data? Are there any **tools** other than interviews that can be used for this purpose?