Name:	Rosa Gini
Affiliation:	Agenzia regionale di sanità della Toscana
Email:	rosa.gini@ars.toscana.it
Presentation type (s):	**Poster**

Obtaining estimates of outcome validity from a small set of parameters: the *component strategy* from the ADVANCE project

Rosa Gini, PhD[1], Caitlin Dodd, MSc[2], Kaatje Bollaerts, PhD[3], Miriam Sturkenboom, PhD[3,4]

[1]Agenzia regionale di sanità della Toscana, Florence, Italy; [2]Erasmus Medical Center, Rotterdam, Italy; [3]P-95, Leuven, Belgium; [4]Utrech Medical Center, Utrecht, The Netherlands

Abstract

Validity indices of the case-finding algorithm of a study outcome can be used to adjust effect estimates by misclassification errors. Obtaining estimates of validity indices is often difficult in multi-database studies. We propose a set of analytical interrelations among validity indices that narrow down the problem to obtaining a small set of input parameters.

Introduction

Misclassification of the study outcome can bias significantly the estimate of the effect of the exposure, and bias can be heterogeneous in multi-database studies, causing observed heterogeneity. Conducting chart based validation studies of case-finding algorithms defining study outcomes is often unfeasible, due to resource limitations or privacy issues. As a consequence, outcome misclassification is not quantified.

In this work we prove that the complete set of validity indices can be obtain from a small set of input parameters.

Interrelations among validity indices of a case-finding algorithm

It is easy to prove from definitions that the following system of 3 equations with 6 parameters holds.

$$\begin{cases} P\,PPV = SE\pi \\ NPV(1-P) = SP(1-\pi) \\ P = SE\pi + (1-SP)(1-\pi) \end{cases}$$

π	true prevalence
P	observed prevalence
SE	sensitivity
SP	specificity
PPV	positive predictive value
NPV	negative predictive value

Since observed prevalence is a parameter that is always known, this implies that from knowledge or any other two parameters the other 3 can be analytically derived by solving the system. We developed a freely available tool that allows to compute the derived indices from any given triplets, as well as uncertainty intervals [1].

Interrelations among validity indices of two component algorithms A and B and their composition A *OR* B

In the European EMIF project a component algorithms strategy was defined, where case-finding algorithms are split in simpler algorithms, each defined by a specific quadruple of data domain (diagnosis, procedures, drug utilization, …), semantics, setting of healthcare where the information was collected (primary care, secondary outpatient care, inpatient care, emergency care, …) [2]. The rationale is that the dimensions of a component algorithm partially explain the validity.

This strategy is used and extended in the ADVANCE European Project[3]. Each database of the network participating in a multi-database study may be able to implement only a subset of the component algorithms, and may define its study outcome as the composition (via a OR logical connector) of a particular subset of components.

Similarly as in the previous section, it can be proven that the validity of the composition is interrelated with the validity of the components. This allows to compute all the indices starting from any combination of 3 parameters between validity indices of the components or of the composite, or true prevalence. For instance if PPV of the components and true prevalence are known, sensitivity and PPV of the composite can be obtained by the following formulas

$$\begin{cases} SE_{A \text{ OR } B} = \dfrac{P_A \ PPV_A}{\pi} + \dfrac{P_B \ PPV_B}{\pi} - \dfrac{P_{A \text{ AND } B} \max(PPV_A, PPV_B)}{\pi} \\ PPV_{A \text{ OR } B} = \dfrac{SE\pi}{P} \end{cases}$$

and NPV and specificity can be obtained using formulas from the previous section.

Application

The problem of assessing validity of case-finding algorithms can be reduced to a small set of input parameters. Input parameters can be estimated from ad-hoc validation studies, or obtained by assuming transportability of parameters found in the literature, or by developing scenarios. The rest of the information is obtained empirically from observing the prevalence of the component algorithms and of their intersections.

Conclusion

This set of formulas may be implemented in the OHDSI set of tools and support exploration of the validity of the case-finding algorithms used to define study outcomes, based on information that can be found in the literature, and on empirical observation.

References

1. P-95. Validity Indices Interrelations. http://apps.p-95.com/Interrelations/
2. Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, et al. Identifying Cases of Type 2 Diabetes in Heterogeneous Data Sources: Strategy from the EMIF Project. PLoS ONE. 2016;11(8):e0160648.
3. ADVANCE project. http://www.advance-vaccines.eu/