APPLICATIONS OF LATENT CLASS ANALYSIS IN DIAGNOSTIC VALIDATION, THE USER PERSPECTIVE

Author's name(s): Boelaert M¹*, Verloo D.², Van der Stuyft P ¹
Affiliation(s): ¹Prince Leopold Institute of Tropical Medicine, Nationalestraat 155, B 2000, Antwerpen, Belgium
²Veterinary and Agrochemical Research Centre, Groeselenberg 99, B 1180 Ukkel, Belgium
Email: boelaert@itg.be
Phone: 0032 3 247 63 05; Fax: 0032 3 247 62 58
Corresponding author: Boelaert M

Keywords: validation studies, gold standard, latent class analysis

Topic area of the submission: Use of Latent Class Models in Medical Diagnosis

Abstract
INTRODUCTION

Diagnostic validation research is often hindered by the lack of a gold standard. The evaluation of new diagnostic tools against reference standards that are “not gold but alloyed” leads to misclassification bias or “reference test bias” [1]. This systematic error can go in either direction, i.e. over- as well as underestimation of the performance of the new tests [1-3]. In the field of tropical medicine, there are several notorious examples of the gold standard problem, a.o. the problematic validation of sleeping sickness, visceral leishmaniasis and tuberculosis diagnostics. Direct microscopic examination of tissue samples or culture can demonstrate the presence of the pathogen without ambiguity, so specificity is close to 100%, but the sensitivity of these procedures is too low to make for a gold standard [4]. Another example is the methodological discussion on the validation of non-culture tests (PCR, LCR) for *Chlamydia* and *N.gonorrhoea* detection [1,5-7]. The reference standard, culture, is again highly specific, but its sensitivity is clearly suboptimal. Staquet el al. showed that if a reference standard is truly 100% specific, but less than 100 % sensitive, the sensitivity estimates of the new test in the classical 2x2 contingency table will be correct, but the specificity estimate will be underestimated [8]. Several researchers in this field have used the method of discrepant analysis, in which discordant results between the new test and the reference test are investigated with a third test. Miller showed mathematically how under
ideal conditions, i.e. when the third test would be 100 % sensitive and 100 % specific, discrepant analysis will nearly always overestimate the performance of the new test [1]. Discrepant analysis violates one of the basic tenets in diagnostic test assessment, saying that the reference test should not depend on the new test under evaluation, and its use was therefore discouraged as an inherently biased method.

In this context, Latent Class Analysis (LCA) was suggested as a potential solution to the gold standard issue [9]. LCA is a mathematical technique that models associations between observed variables that imperfectly measure a non-observable (latent) variable [10,11]. In basic latent class models, the observed variables are said to be conditionally independent (conditional on latent class), i.e. there are no associations between the observed variables within each category of the latent variable. More advanced models exist, where this condition is relaxed [9,12]. When a diagnostic test is evaluated on a group of study subjects, their true disease status can be considered as a latent variable with two mutually exclusive and exhaustive classes or categories, "diseased" and "non diseased". Given a group of individuals with unknown disease status, for whom results from several diagnostic tests are available, LCA will model the probability of each combination of test results conditional on latent class, i.e. disease status. The LCA model hence produces an estimate of disease prevalence and of sensitivity and specificity of all the diagnostic tests. Inference can be made in a frequentist or a Bayesian framework.

We explored the value of LCA by comparing it with the classical validation approach in a number of validation studies we undertook since 1999, and present some of the strengths and weaknesses of the method from the perspective of the user.

METHODS

We used the following data sets:

1. Dogs suspected of infection with L. infantum (n=152), sera collected in Tunisia. The manifest variables recorded were a clinical case definition, DME of tissue aspirate, parasitological culture, and three serological tests: IFAT, ELISA, and DAT. [13]
2. Clinical suspects of visceral leishmaniasis disease (n=145), recruited in an endemic area in Sudan. Observed diagnostic test variables were IFAT, parasitology and DAT and , to control for possible confounding, three observed external variables: prior treatment, sex and age group were included in the models [14].
3. Sera of Bolivian blood donors (n =400) screened for T. cruzi infection with 7 serological tests: 2 IHA, 1 IFAT and 4 ELISA tests [15].
4. Clinical suspects of visceral leishmaniasis \( (n = 310) \) recruited in a tertiary care setting in Nepal. One sign, pancytopenia, and 5 diagnostic tests were compared: parasitology, the formol-gel test, IFAT, a rK39 dipstick, and DAT [16].

5. Patients suspected of tuberculosis \( (n = 300) \) recruited in a tertiary care setting in Rwanda. Fever, hemoptysis, DME, radiological signs and culture were used as manifest variables.

For each data set, validation of the diagnostic tests was first attempted by the classical method: comparison of each single test with the classification by a reference standard in the 2x2 contingency table. Subsequently, we fitted a series of LCA models to the data. In addition we explored the use of Bayesian estimation methods as an alternative estimation method.

**DISCUSSION**

LCA is conceptually more satisfactory than discrepant analysis to address the issue of the failing reference standard. Loglinear latent class modeling moreover provides a flexible approach and offers the possibility to control for confounding and to take missing data patterns into account. We confronted LCA with classic validation in five data sets. In accordance with theoretical expectations, LCA corroborated the sensitivity estimates of the classic approach in studies 1, 2, and 4 and produced higher specificity estimates for serological tests than those obtained by the classical approach. In this way, LCA eluded the inherent bias in the classical approach. However, a certain number of technical problems arose related to the estimation methods, as special caution is needed in the case of sparse data sets [17]. Some of these problems could be circumvented by a Bayesian approach. For identifiable models and when no prior information is included in the Bayesian model both the frequentist and the Bayesian approaches yielded similar results. However, in the Bayesian framework, posterior distributions of the parameters, mostly obtained by Markov Chain Monte Carlo techniques are very flexible for further simulation of other parameters of interest. Besides this, normal distribution approximations, in the frequentist framework commonly employed to derive confidence intervals around unknown parameters from estimated standard errors, are not required.

In summary, LCA is a valuable methodology for diagnostic validation studies, though less intuitive than the classic approach in the 2 x 2 table. LCA should complement the classical approach whenever no gold standard is available.
References


