Latent Class Analysis to Determine the Accuracy of Diagnostic Tests in Orthopaedics

Geert A. Buijze, MD[1], Timothy E. Hanson, PhD[2], Wesley O. Johnson, PhD[3], David Ring MD, PhD[1]
[1]Massachusetts General Hospital, [2]University of Minnesota, [3]University of California

INTRODUCTION

The scientific study of diagnostic tests is often hindered by the lack of a consensus reference standard. Inadequate reference standards may account for much of the heterogeneity in studies of diagnostic performance characteristics.

Latent class analysis is a statistical technique that provides estimates of diagnostic performance characteristics in the absence of a consensus reference standard(1). The method relies on having multivariate categorical data based on multiple diagnostic test results. The estimation of test accuracies and prevalence are performed with use of the maximum likelihood or Bayesian methodologies, or both. In plain speak, latent class analysis looks for groups of test results (or latent classes) that represent levels of disease probability. Latent class analysis has been used in many diagnostic and prevalence studies of diseases lacking a consensus reference standard, including various infectious diseases(2-5), psoriatic arthritis(6), and carpal tunnel syndrome(7).

The realization that we lack—and may always lack—consensus reference standards for some diseases represents a paradigm shift that is somewhat counterintuitive. We tend to think of diseases as "all or none." Either our diagnosis is correct, or it's incorrect. Latent class analysis is another approach to medical science that emphasizes that we are often dealing with probabilities of diagnosis rather then true knowledge of the diagnosis. Adopting a mindset that we are giving our patients estimates of how likely certain diagnoses exist and therefore how likely certain treatments are going to be necessary or helpful may help us better manage the large proportion of diseases that are inherently uncertain.

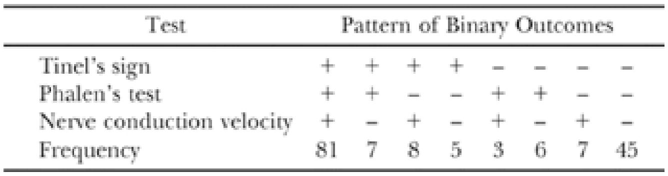
The aim of this paper is to explain the methodology of latent class analysis and to discuss the possibilities for its use in orthopedic diagnostic studies.

STATISTICS

Latent class analysis requires a set of results from three or more tests in a selected cohort. Depending on whether the results of the tests are related, two methods can be used.

The maximum likelihood (ML) based method developed by Hui and Walter in 1980(2) results in estimates of sensitivity and specificity of each test and prevalence of the disease in the study population, under the assumption of conditional independence of the tests. Since its development, latent class analysis has been modified to be applicable in a great variety of studies. Walter designed the program LATENT1 (Latent1 Software, Version 3, McMaster University, Hamilton, Ontario, Canada), which calculates the maximum likelihood estimates and gives confidence intervals for test accuracies and prevalence. In addition to the basic parameters, LATENT1 provides positive predictive values for each pattern of test results.

In contrast, when test results are expected to correlate, the data will violate the conditional independence assumption of standard latent class analysis. In that case, a recently developed latent class model can be used, which allows for conditional dependence among multiple test results(8, 9). These methods rely on Bayesian statistical methodology rather than on maximum likelihood, and they generalize the Bayesian version of the Hui and Walter model which was developed by Johnson, Gastwirth and Pearson in 2001(15). If the sample size is relatively small for the number of tests performed in a study, the model may require incorporation of some expert input as guidance in the form of prior distributions.



Test	Pattern of Binary Outcomes							
Tinel's sign	+	+	+	+	–	–	–	–
Phalen's test	+	+	–	–	+	+	–	–
Nerve conduction velocity	+	–	+	–	+	–	+	–
Frequency	81	7	8	5	3	6	7	45

Table 1. Frequency of test results observed in 81 participants (162 wrists)(7). (Reprinted from LaJoie AS, McCabe SJ, Thomas B, Edgell SE. Determining the sensitivity and specificity of common diagnostic tests for carpal tunnel syndrome using latent class analysis. Plast Reconstr Surg. 2005 Aug;116(2):502-7.)

[1]Orthopaedic Hand and Upper Extremity Service
Massachusetts General Hospital, Harvard Medical School
Yawkey Center, Suite 2100, 55 Fruit Street
Boston, MA 02114, USA

[2]Division of Biostatistics
University of Minnesota, A460 Mayo Building
Minneapolis, MN 55455, USA

[3]Department of Statistics
University of California, 2232 Bren Hall
Irvine, CA 92697, USA

Corresponding author:

David Ring, MD, PhD
Massachusetts General Hospital, Harvard Medical School
Yawkey Center, Suite 2100, 55 Fruit Street
Boston, MA 02114, USA
Tel: +1 617 643 1267
Fax: +1 617 726 0460
dring@partners.org

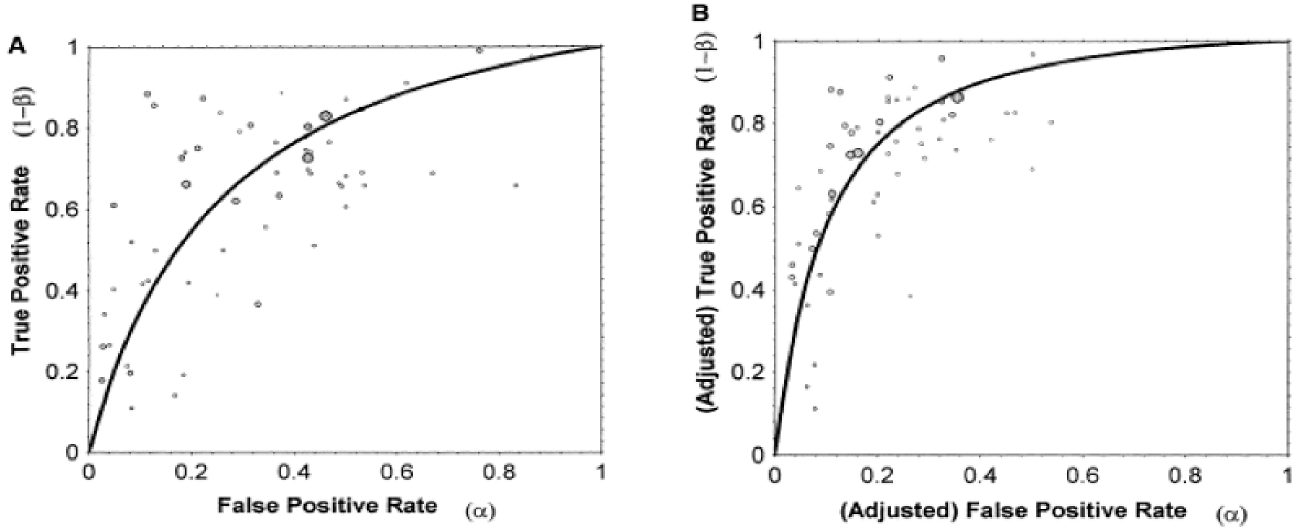


Figure 1.
True-positive rate vs. false-positive rate, with fitted SROC curves: Pap smears as a test of cervical histological abnormality. (A) With observed data; (B) with data adjusted for misclassification of the reference standard. Each point represents one study with area proportional to total sample size(10). (Reprinted from Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol. 1999 Oct;52(10):943-51.)

Meta-analyses of diagnostic tests can also account for the lack of a reference standard by calculating adjusted SROC curves using pooled diagnostic values—an approach that allows for the possibility of errors in the reference standard, through use of a latent class model(10). The model presumes that the true disease status of each subject is unknown, or latent, and uses parameter estimates to calculate a set of fitted frequencies for the numbers of true (but unobserved) cases and noncases, adjusted for the misclassification in the reference standard.

Example 1: Diagnosis of Carpal Tunnel Syndrome

There is no consensus standard for defining the presence or absence of carpal tunnel syndrome(11). LaJoie et al. applied latent class analysis to quantify the accuracy rates of the nerve conduction velocity test, Tinel's sign, and Phalen's test (7). All three tests were performed on 162 wrists in 81 patients.

The most observed frequency of test results was a positive result in all three measures and the second most observed frequency was a combination of negative results in all (Table 1). Latent class analysis using LATENT1 software estimated a prevalence of 60% in the sample population. Tinel's sign had the highest sensitivity and specificity of the three tests. Compared to previous studies using nerve conduction velocity as reference standard, latent class analysis resulted in higher estimates for the sensitivity and specificity of Tinel's sign and Phalen's test. This is consistent with the fact that the use of an imperfect reference standard may result in underestimation of the accuracy of diagnostic tests.

Example 2: Diagnosis of Scaphoid Fractures

There is no reference standard for the diagnosis of a true fracture among patients with suspected scaphoid fracture, and it is likely that there never will be. We applied latent class analysis to data from two prospective cohort studies on patients with suspected scaphoid fractures: one trial compared MRI to CT 34 patients (Cohort 1) and the other compared MRI to bone scintigraphy and clinical tests (Cohort 2) in 78 patients. As we assumed that the diagnostic tests of Cohort 1 (MRI versus CT) met the conditional independence criteria, we used LATENT1 software for the latent class analysis. In Cohort 2, we used a Bayesian analysis which allowed for conditional dependence between the 7 clinical tests. Calculations using reference standards and latent class analysis were compared.

In the first cohort, the sensitivity and specificity in the latent class analysis were slightly higher than the calculations made using a reference standard for both the CT in the scaphoid planes and the MRI. In the second cohort, the biggest differences were in the sensitivity of MRI (9% higher with latent class analysis) and the sensitivities of physical examination maneuvers, which were between 27% and 37% lower with latent class analysis. The results of the latent class analysis reflected the imperfections in the reference standards used in both studies.

Example 3: Diagnoses in Psychiatry

A field that is notorious for latent classes (or unobservable states) of illness is psychiatry(12). Simple examples of latent classes from everyday life are seen in the human tendency to classify people according to personality attributes (honest or dishonest), emotional states (happy or sad), or intellectual ability (intelligent or unintelligent). Honesty, happiness, or intelligence, cannot be observed directly but the behavior from which these latent (i.e., unobservable) characteristics are inferred can be observed. Since no consensus standard can be established for psychiatric diagnoses, one must rely on measures of observer agreement to quantify the quality of diagnostic judgments.

To demonstrate the utility of consistency tests, Faraone and Tsuang(12) analyzed data from a preexisting data set on diagnosis of major depressive disorder(13). The original presentation of these data used a logistic regression model of diagnostic agreement to compute diagnostic performance statistics for major depression. The authors calculated parameter estimates for diagnostic accuracy according to the latent class model and compared it to the parameters from the logistic regression. The two analyses modeled different phenomena and used overlapping but different information. The results of both statistical methods were highly consistent. Both found high diagnostic accuracy for depression, and both concluded that specificity was greater than sensitivity.

Example 4: Latent Class Analysis in Meta-Analyses

Meta-analysis of diagnostic data entails more than simply pooling all the data into a fourfold table as simple pooling may cause serious bias because of confounding of disease prevalence and test thresholds used in the contributing studies(14). As stated in the Statistics section of this paper, Walter et al. proposed a method that has the primary goal of estimating diagnostic accuracy parameters using the SROC curve while taking errors in the reference standard into account(10).

They used preexisting data set from a meta-analysis of 59 studies on the performance of the Pap smear for the diagnosis of cervical precancer. The studies showed a wide variation in accuracy parameters and were highly negatively correlated, indicating important differences in threshold between studies. Based on overall parameter estimates, data frequencies of primary studies are revised by calculating fitted data frequencies, and the estimates of sensitivity and specificity are updated for each study. Compared to the original data, adjustment for the reference standard errors has reduced the scatter of points in the scatter plots, and narrowed the range, particularly of the false-positive rates (Figure 1).

CONCLUSION

The diagnostic performance characteristics calculated using latent class analysis are often different from those calculated according to standard formulas based on a reference standard. Because these differences may reflect shortcomings of the reference standard, we recommend that latent class analysis be considered when calculating diagnostic performance characteristics for any disease that lacks a reliable or consensus reference standard. Given the inherent uncertainty in many diagnostic tests it may be appropriate—for many if not most illnesses—that patients and doctors base decisions on probabilities of disease rather than the traditional dichotomous, all or none, concept of disease.

Acknowledgements:

GAB would like to thank the Netherlands Organization for Scientific Research (NWO) for their support to this research.

References

1. **Hui SL, Walter SD.** Estimating the error rates of diagnostic tests. Biometrics. 1980 Mar;36(1):167-71.
2. **Baughman AL, Bisgard KM, Cortese MM, Thompson WW, Sanden GN, Strebel PM.** Utility of composite reference standards and latent class analysis in evaluating the clinical accuracy of diagnostic tests for pertussis. Clin Vaccine Immunol. 2008 Jan;15(1):106-14.
3. **De La Rosa GD, Valencia ML, Arango CM, Gomez CI, Garcia A, Ospina S, et al.** Toward an operative diagnosis in sepsis: a latent class approach. BMC Infect Dis. 2008;8:18.
4. **Butler JC, Bosshardt SC, Phelan M, Moroney SM, Tondella ML, Farley MM, et al.** Classical and latent class analysis evaluation of sputum polymerase chain reaction and urine antigen testing for diagnosis of pneumococcal pneumonia in adults. J Infect Dis. 2003 May 1;187(9):1416-23.
5. **Tuyisenge L, Ndimubanzi CP, Ndayisaba G, Muganga N, Menten J, Boelaert M, et al.** Evaluation of latent class analysis and decision thresholds to guide the diagnosis of pediatric tuberculosis in a Rwandan reference hospital. Pediatr Infect Dis J. 2010 Feb;29(2):e11-8.
6. **Symmons DP, Lunt M, Watkins G, Helliwell P, Jones S, McHugh N, et al.** Developing classification criteria for peripheral joint psoriatic arthritis. Step I. Establishing whether the rheumatologist's opinion on the diagnosis can be used as the «gold standard». J Rheumatol. 2006 Mar;33(3):552-7.
7. **LaJoie AS, McCabe SJ, Thomas B, Edgell SE.** Determining the sensitivity and specificity of common diagnostic tests for carpal tunnel syndrome using latent class analysis. Plast Reconstr Surg. 2005 Aug;116(2):502-7.
8. **Dendukuri N, Joseph L.** Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001 Mar;57(1):158-67.
9. **Qu Y, Tan M, Kutner MH.** Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. Biometrics. 1996 Sep;52(3):797-810.
10. **Walter SD, Irwig L, Glasziou PP.** Meta-analysis of diagnostic tests with imperfect reference standards. J Clin Epidemiol. 1999 Oct;52(10):943-51.
11. **Rempel D, Evanoff B, Amadio PC, de Krom M, Franklin G, Franzblau A, et al.** Consensus criteria for the classification of carpal tunnel syndrome in epidemiologic studies. Am J Public Health. 1998 Oct;88(10):1447-51.
12. **Faraone SV, Tsuang MT.** Measuring diagnostic accuracy in the absence of a «gold standard». Am J Psychiatry. 1994 May;151(5):650-7.
13. **Rice JP, Endicott J, Knesevich MA, Rochberg N.** The estimation of diagnostic sensitivity using stability data: an application to major depressive disorder. J Psychiatr Res. 1987;21(4):337-45.
14. **Irwig L, Macaskill P, Glasziou P, Fahey M.** Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol. 1995 Jan;48(1):119-30; discussion 31-2.