行政院國家科學委員會專題研究計畫　成果報告

潛在變數模型在評估診斷檢驗上的應用及其軟體的發展

(2/2)

計畫類別：個別型計畫
計畫編號：NSC93-2118-M-009-005-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：國立交通大學統計學研究所

計畫主持人：黃冠華

報告類型：完整報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 10 月 31 日

行政院國家科學委員會補助專題研究計畫成果報告

潛在變數模型在評估診斷檢驗上的應用及其軟體的發展(1/2)(2/2)

計畫類別：個別型計畫
計畫編號：NSC 92-2119-M-009-002; NSC 93-2118-M-009-005
執行期間：92 年 12 月 01 日至 94 年 07 月 31 日

計畫主持人：國立交通大學統計學研究所黃冠華教授
共同主持人：無
計畫參與人員：吳進結、陳穗碧、高徽宜

成果報告類型(依經費核定清單規定繳交)：完整報告

本成果報告包括以下應繳交之附件：無

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

執行單位：國立交通大學統計學研究所

中　華　民　國　九十四　年　十　月　三十一　日

目　　錄

一、 中英文摘要

中文摘要

在許多醫學研究中，我們常因為花費，時間或無法量度等的問題而觀察不到最想要的觀察值。此時，一個有效的替代變數便常被用來取代這個我們所觀察不到的結果。例如，生物標記常被用來代替觀察到真實的癌症例子，這是因為我們通常需要很長的時間才能觀察到癌症的例子。 我們常用一系列的問卷問題來描述老年人之功能障礙，這是因為我們無法找到一單獨量度值來表示功能障礙。

潛在變數模型被認為是一個分析替代變數最有效的統計方法。這個模型先用觀察到的替代變數來估計我們看不到的結果， 然後再去求我們估計出來的結果和危險因子間的關係。這個模型具有簡潔和可辨認量度誤差的特性，且此模型能得到我們真正有興趣的結果和危險因子間的關係。

本計畫主持人最近在一篇論文中提出了一個非常有彈性的潛在變數模型。我也在另外一篇論文中發表如何選取潛在變數之維數。 在這個計畫中，我將應用潛在變數模型來評估診斷檢驗之有效性和可靠性。有了此一方法，我們將可去作診斷評估而不須知道病人的真正疾病狀態。除此之外，我也將要把所有我所提出分析潛在變數的方法統合起來成一電腦軟體，且此一軟體將可很簡易的被所有的人來使用。

關鍵字: 潛在變數，可靠性，電腦軟體，替代變數，有效性，網頁。

Abstract

In many medical studies, the definitive outcome is inaccessible due to cost, time and difficulty of measurement. A valid surrogate endpoint is then measured in place of the biologically definitive or clinically most meaningful endpoint. For example, biomarkers are very often used as surrogates of observing new cases of cancer in testing treatments for cancer prevention, where event rates are low and a long time may be needed to obtain definitive outcomes. Functional disability is commonly quantified as self-reported responses to a series of questions about difficulty performing tasks of routine living, because no obvious single measure of disability exists.

Latent variable models are recognized as an effective statistical tool for analyzing surrogates. Such models treat the unobserved definitive outcome as the one to be analyzed for relations with risk factors, and the measured surrogates as quantities that imperfectly determine the object of interest. Analysis based on latent variable models is parsimonious, explicitly recognize errors in measurement, and can give well-summarized inferences on the theory underlying the choice of

surrogates.

The author has recently developed a very flexible latent variable model for analyzing surrogates. I also addressed issues about selecting the number of dimensions needed to characterize the surrogates' distribution. In this project, I propose to extend the latent variable model developed by the author to evaluate the validity and reliability of diagnostic tests. This application is important because the model enables us to evaluate tests without knowing the gold standard. I also plan to incorporate the latent variable methodologies developed by the author into an easy-to-use statistical software capable of being used effectively by all levels of participants of data analysis.

Keywords: latent variable, reliability, software, surrogate endpoint, validity, web page.

二、 緣由與目的

In many medical studies, the definitive outcome is inaccessible due to cost, time and difficulty of measurement. A valid surrogate endpoint is then measured in place of the biologically definitive or clinically most meaningful endpoint. For example, biomarkers are very often used as surrogates of observing new cases of cancer in testing treatments for cancer prevention, where event rates are low and a long time may be needed to obtain definitive outcomes (Piantadosi 1997). Functional disability is commonly quantified as self-reported responses to a series of questions about difficulty performing tasks of routine living (e.g., Katz et al. 1963), because no obvious single measure of disability exists.

Models that permit exploration of relationships between unobservable variables and their surrogates are referred to as latent variable models. My research is focused on analyzing data collected in situations where investigators use multiple discrete indicators as surrogates, for example, a set of questionnaires. These models are called the latent class analysis (LCA) model (Lazarsfeld and Henry 1968, Goodman 1974). The basic LCA postulates an underlying categorical latent variable, and this latent variable can explain the association seen among measured items. Recently, several authors extended the LCA model to incorporate covariate effects on estimating the underlying mechanism (Dayton and Macready 1988, Bandeen-Roche et al. 1997), or on estimating measured indicator distributions within latent classes (Melton et al. 1994). I and co-author proposed an LCA model that used covariates on describing distributions of both the underlying latent class and the measured indicators themselves (Huang and Bandeen-Roche 2003A). Significantly, we developed a model framework that guarantees identifiability of the two types of covariate effects.

In this project, I propose to extend the latent class model developed by the author to evaluate the validity and reliability of diagnostic tests. This application is important because the model enables us to evaluate tests without knowing the gold standard. I also plan to incorporate all the latent variable methodologies developed by the author into an easy-to-use statistical software

capable of being used effectively by all levels of participants of data analysis. These methodologies include the proposed LCA mode (Huang and Bandeen-Roche 2004), and selecting the number of dimensions needed to characterize the surrogates' distribution (Huang 2005).

三、 結果與討論

In the following, I will discuss our results in each aim separately.

Aim 1: To apply the developed latent class model in evaluating the validity and reliability of diagnostic tests while no golden standard exists.

When evaluating and comparing diagnostic tests for a certain disease, it is often that none of the tests can be considered as a gold standard and be used to represent an individual's true disease status. As a result, tests' sensitivities, specificities and Kappa statistics, which are the most commonly used indices to determine the validity and reliability of a test, cannot be obtained. In fact, one may argue that this is virtually always the case, since few tests are considered to be 100% accurate. Despite these limitations, it is important for clinical and public health practices to develop methods that best estimate these two indices.

Because of the ability in estimating the underlying definite outcome, the latent class model is a perfect tool for evaluating the validity and reliability of diagnostic tests while no golden standard exists. Traditional latent class models assume that test results are independent, conditional on an individual's true disease status. In reality, however, the diagnostic tests are often correlated within the diseased and the disease-free population (Qu, Tan, Kutner 1996). When correlations between tests are ignored, the comparisons in sensitivity and specificity between tests are inefficient. The purpose of this project is to present a latent class model not based on the above conditional independence assumption.

Many studies have tried to address the problem of evaluating tests in the absence of a gold standard. Qu, Tan and Kutner (1996) proposed a methodology that allows for dependence in the repeated tests conditional on the true status by incorporating an individual-specific Gaussian random effect in the probability distribution that links the observed test results to the true status. Uebersax and Grove (1993) described a model related to Qu et al.'s (1996) that assumes a latent trait distributed as a mixture of two Gaussian distributions. Albert et al. (2001) developed a mixture model that describes between-specimen heterogeneity that they expect in their bladder tumor data.

Despite this body of research, there are three important issues remain unsolved by the literature. First, previous literature assumed Gaussian distributions for random effects or for the underlying latent variable. Inferences of sensitivity and specificity may be highly sensitive to this assumption (Albert et al. 2001). Second, while the validity of a test, measured by sensitivity and specificity,

can be obtained from the previous models, a test's reliability, measured by Kappa statistic, does not come from these models. It is known that both validity and reliability are important for evaluating diagnostic tests. An approach that can estimate validity and reliability simultaneously is extremely valuable. Third, the latent class model is an excellent approach for evaluating tests. However, the model is often very complicated, thus model identifiability and assumptions can not be guaranteed. Methods for checking model identifiability and various assumptions are needed.

We have developed a framework for simultaneously estimating a test's sensitivity, specificity and Kappa statistic. This approach applies latent class models and does not assume a Gaussian distribution. More specifically, suppose Y_{ij} is the test result for individual i at test j and the possible values of Y_{ij} are 0= test negative and 1=test positive. S_i represents the true disease status for individual i with values 1=having disease and 2=disease free. ($\boldsymbol{x_i}$, $\boldsymbol{z_i}$, $\boldsymbol{h_i}$) are the covariates associated with the ith individual. If we assume the "second-order dependence", then the likelihood of Y_{ij}'s can be expressed as

$$\begin{aligned}
&\Pr(Y_{i1} = y_1, \ldots, Y_{iM} = y_M \mid \boldsymbol{x_i}, \boldsymbol{z_i}, \boldsymbol{h_i}) \\
&= \prod_{j=1}^{2} \Pr(S_i = j \mid \boldsymbol{x_i}) \prod_{m=1}^{M} \Pr(Y_{im} = y_m \mid \boldsymbol{z_{imj}}, S_i = j) \times \\
&\left\{ 1 + \sum_{m<v} \frac{(y_m - p_{imj})(y_v - p_{ivj})}{p_{imj}(1-p_{imj})\, p_{ivj}(1-p_{ivj})} \left[\Pr(Y_{im} = 1, Y_{iv} = 1 \mid \boldsymbol{h_{imvj}}, S_i = j) - p_{imj}\, p_{ivj} \right] \right\},
\end{aligned}$$

where $p_{imj} = \Pr(Y_{im}=1 \mid \boldsymbol{z_{imj}}, S_i=j)$. We then define $\Pr(S_i=1 \mid \boldsymbol{x_i})$, $\Pr(Y_{im}=1 \mid \boldsymbol{z_{imj}}, S_i=j)$ and $\Pr(Y_{im}=1, Y_{iv}=1 \mid \boldsymbol{h_{imvj}}, S_i=j)$ as follows:

$$\begin{aligned}
\log\left[\frac{\Pr(S_i = 1 \mid \boldsymbol{x_i})}{\Pr(S_i = 2 \mid \boldsymbol{x_i})} \right] &= \boldsymbol{x_i \beta} \\
\log\left[\frac{\Pr(Y_{im} = 1 \mid \boldsymbol{z_{imj}}, S_i = j)}{\Pr(Y_{im} = 0 \mid \boldsymbol{z_{imj}}, S_i = j)} \right] &= \boldsymbol{z_{imj} \alpha} \\
\log\left[\frac{1 + \kappa_{imvj}}{1 - \kappa_{imvj}} \right] &= \boldsymbol{h_{imvj} \tau},
\end{aligned}$$

where κ_{imvj} is the Kappa statistic of Y_{im} and Y_{iv}, and is a function of $\Pr(Y_{im}=1, Y_{iv}=1 \mid \boldsymbol{h_i}, S_i=j)$. Under the proposed model, a test's sensitivity, specificity and Kappa statistic can be obtained in terms of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$.

We use generalized estimating equation (GEE) (Liang and Zeger 1986) approach to estimate the parameters. To estimate $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$, the following estimating equation is used

$$U_{\varphi} = \sum_{i=1}^{N} \left(\frac{\partial E(Y_i)}{\partial \varphi} \right)^T (Var(Y_i))^{-1} [Y_i - E(Y_i)],$$

where $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{iM})$. To estimate $\boldsymbol{\tau}$, the following estimating equation is used

$$U_{\tau} = \sum_{i=1}^{N} \left(\frac{\partial E(r_i)}{\partial \tau} \right)^T (W_i(\varphi, \tau))^{-1} [r_i - E(r_i)],$$

where

$$r_i = \left\{ [Y_{i1} - E(Y_{i1})][Y_{i2} - E(Y_{i2})], [Y_{i1} - E(Y_{i1})][Y_{i3} - E(Y_{i3})], \cdots, [Y_{i(M-1)} - E(Y_{i(M-1)})][Y_{iM} - E(Y_{iM})] \right\}$$

is a $M(M-1)/2 \times 1$ vector; and

$$W_i(\varphi, \tau) = B_i^{1/2} R_i(\varphi, \tau) B_i^{1/2},$$

where B_i is a $M(M-1)/2 \times M(M-1)/2$ diagonal matrix with the (*m*, *v*)'th element being $Var\{[Y_{im} - E(Y_{im})][Y_{iv} - E(Y_{iv})]\}$; and $R_i(\varphi, \tau)$ is a $M(M-1)/2 \times M(M-1)/2$ working correlation matrix. One can use Fisher scoring to obtain the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$. It can be shown that the estimator $(\hat{\beta}, \hat{\alpha}, \hat{\tau})$ is asymptotically Gaussian with asymptotic variance

$$AV = \left(\sum_{i=1}^{N} E_i^T G_i^{-1} F_i \right)^{-1} \left(\sum_{i=1}^{N} E_i^T G_i^{-1} Cov(Y_i, r_i) G_i^{-1} E_i \right) \left(\sum_{i=1}^{N} F_i^T G_i^{-1} E_i \right)^{-1},$$

where

$$E_i = \begin{bmatrix} \partial E(Y_i)/\partial \varphi & 0 \\ 0 & \partial E(r_i)/\partial \tau \end{bmatrix}, F_i = \begin{bmatrix} \partial E(Y_i)/\partial \varphi & 0 \\ \partial E(r_i)/\partial \varphi & \partial E(r_i)/\partial \tau \end{bmatrix}, G_i = \begin{bmatrix} Var(Y_i) & 0 \\ 0 & W_i(\varphi, \tau) \end{bmatrix},$$

and $Cov(Y_i, r_i)$ can be estimated by

$$\begin{bmatrix} Y_i - E(Y_i) \\ r_i - E(r_i) \end{bmatrix} \begin{bmatrix} Y_i - E(Y_i) \\ r_i - E(r_i) \end{bmatrix}^T.$$

The proposed model is applied to determine the minimum number of eye measurements which must be made to achieve a "good enough" arteriole-venule ratio (AVR). AVR expresses the relationship of the diameters of the eye's arteries and veins, and is thereby an indicator of elevated blood pressure which in turn indicates heart disease. The AVR data is from the Arteriosclerosis Risk in Communities Study (ARIC), which is an epidemiological survey of risk factors for heart disease in 15,792 older adults in communities in four states: North Carolina, Mississippi, Minnesota, and Maryland (Hubbard et al. 1999).

The AVR has been shown to be associated with elevated blood pressure in the ARIC population, while arterioles alone have only a weak association (Hubbard et al. 1999). Additionally, using a ratio reduces the effect of measurement error, assuming that the error is the same for both vessel types. In the ARIC study, different graders measured each several of the diameters of each participant's arterioles and venules. Various vessel diameters, from large to small, were measured. Minimizing the number of arterioles and venules which must be measured is desirable because it is difficult and very costly to measure small vessels. Measurements may contain substantial errors, so multiple measurements must be made despite these obstacles. Therefore, we seek a statistical method to determine a sufficient number of vessel measurements for the purpose of using AVR as an indicator of elevated blood pressure.

There are two analytic challenges to overcome before determining the appropriate number of vessel measurements. First, we need criteria for a "good" indicator of blood pressure. Second, we need to determine a gold standard in order to determine whether our chosen number of vessel measurements meet our chosen criteria for "close enough" to the gold standard.

Our proposed latent class model is a good candidate for determining the appropriate number of vessel measurement to use for calculating arteriole-venule ratios. A latent variable model is used to summarize AVR measurements and estimate an underlying gold standard. We extend the traditional model to allow conditional dependence in order to overcome the possible assumption violation described above. Sensitivity, specificity, and the Kappa coefficient are obtained from the proposed model, to be used a criteria for judging whether AVR is a good indicator of high blood pressure.

The paper describing above results is in preparation.

Aim 2: To document and incorporate developed methodologies on latent class modeling into friendly and portable software.

We had developed an R package: RLCA to implement our works on the latent variable modeling. The package can fit the regression extension of latent class analysis model (Huang and Bandeen-Roche 2004), and identify the number of dimensions for classifying the observed

population under the regression extension of latent class analysis model (Huang 2005). The RLCA package can (1) read in the raw data and transform them into a ready-to-be-analyzed format, and (2) access and implement developed methods for analysis. A comprehensive user's guide is also contained in the RLCA package. This user's guide (1) introduces the fundamental ideas of latent variable models, (2) works the users through an elementary example of using the developed software, and (3) details the interpretation of analysis results. The RLCA package is posted on the web page:
http://www.stat.nctu.edu.tw/subhtml/source/teachers/ghuang/software.htm.

Aim 3: To create a web page as a resource for researchers interested in latent variable analysis.

We have developed a web page with two main missions: first, to introduce the latent variable model to medical researchers and data analysts, and, second, to share the developed results with the latent variable research community. Therefore, the web page contains (1) comprehensive introduction of the latent variable model, (2) latent variable bibliography, (3) available software, including developed by the author, free shared and commercial products, (4) recent publications from the author and other researchers, and (5) linkage to other related web pages. This web page is located in http://ourworld.compuserve.com/homepages/jsuebersax/index.htm and http://www.stat.nctu.edu.tw/subhtml/source/teachers/ghuang/research.htm.

四、 計畫成果自評

Surrogate endpoints are very popular in the medical research. They are less expensive, easier to measure, and can catch some unobservable concepts. However, analytic results may be biased and lose lots of useful information if appropriate statistical methods are not used. The latent variable model is the statistical methods developed for analyzing surrogates. In the statistical and social science research, many important advances on latent variable modeling have been made in the past two decades. However, these models are not accessible to a broad audience of medical scientists who are impeded by the rigorous mathematics or unavailability of software. This project is trying to fill up this gap by providing a comprehensive web page and an unified software of recently developed latent variable methodologies

In this project, first, we have developed to a very powerful model in evaluating the validity and reliability of diagnostic tests while no golden standard exists. Second, a comprehensive web page is created to introduce the latent variable model and demonstrate the author's ability on independent research. As a result, the latent variable model will receive more attentions from medical researchers and more collaborations and research funding can be expected. Third, an user-friendly latent variable software will be available via the Internet, which provides researchers a convenient way to implement the model. Forth, the principal investigator of this project has applied the developed model and software to several large epidemiology studies and

has lead to international collaboration and produce large impacts on medical research. Fifth, research assistants in the project have had a systematic understanding in latent variable modeling, and have gained experiences in computer programming, web development and real data analysis. Sixth,

We have met all proposed aims set in the proposal. We will finalize our manuscript on evaluating diagnostic tests without golden standard and submit it to the SCI journal. The developed RLCA package is now ready for download on the web page. We believe that the creation and sharing of the RLCA software will greatly increase the impacts of our research on latent class modeling.

五、 參考文獻

Albert PS, McShane LM, Shih JH (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics* 57:610-619.

Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association* 92:1375-1386.

Dayton CM, Macready GB (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association* 83:173-178.

Goodman LA (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61:215-231.

Huang GH (2005). Selecting the number of classes under latent class regression: a factor analytic analogue. *Psychometrika* 70:325-345.

Huang GH, Bandeen-Roche K (2004). Building an identifiable latent variable model with covariate effects on underlying and measured variables. *Psychometrika* 69:5-32.

Hubbard LD, Rosemary JB, King WN, Clegg LX, Klein R, Cooper LS, Sharrett AR, Davis MD, Cai J, ARIC Group (1999). Methods for Evaluation of Retinal Microvascular Abnormalities Associated with Hypertension/Sclerosis in the Arteriosclerosis Risk in Communities Study. *Ophthalmology* 106:2269-2280.

Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffer MW (1963). Studies of illness in the age. The index of ADL: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 185:914-918.

Lazarsfeld PF, Henry NW (1968). *Latent Structure Analysis*. New York: Houghton-Mifflin.

Liang KY and Zeger SL (1986). Longitudinal data and analysis using generalized linear models. *Biometrika* 73: 13-22.

Melton B, Liang KY, Pulver AE (1994). Extended latent class approach to the study of familial/sporadic forms of a disease: its application to the study of the heterogeneity of schizophrenia. *Genetic Epidemiology* 11:311-327.

Piantadosi S (1997). *Clinical Trials: A Methodologic Perspective*. New York: Wiley.

Qu Y, Tan M, Kutner MH (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 52:797-810.

Rubin GS, Bandeen-Roche K, **Huang GH**, Munoz B, Schein OD, Fried LP, West SK (2001). The association of multiple visual impairments with self-reported visual disability: SEE project. *Investigative Ophthalmology and Visual Science* 42:64-72.

Uebersax JS, Grove WM (1993). A latent trait finite mixture model for the analysis of rating agreement. *Biometrics* 49:823-835.