

Application of Latent Curve Models in Medical Research – A Review

Jun-Hao Pan,¹ Xin-Yuan Song² and Timothy Kwok³

1. Lecturer, Department of Psychology, Sun Yat-sen University, Guangzhou; 2. Associate Professor, Department of Statistics, Chinese University of Hong Kong;

3. Professor, Department of Medicine and Therapeutics, Chinese University of Hong Kong

DOI:10.17925/ENR.2009.04.02.52

Abstract

Recently, latent curve modelling (LCM) has received increasing attention in the analysis of longitudinal data. It is a method to model individual change and to assess the effects of co-variables and the relationship among multiple outcomes. It provides an integrated and flexible approach in modelling developmental processes from both inter- and intra-individual perspectives. Similar to conventional longitudinal analysis, the main objectives of this model are to characterise changes in the response of interest over time and to examine the selected covariates that contribute to those changes. In this article the fundamental principle of LCM is briefly introduced. Several important kinds of LCM, including linear LCM, non-linear LCM, multilevel LCM and mixture LCM, together with their applications in medical research, are reviewed. We believe that this statistical technique should become more popular in medical applications, and that the medical field would benefit from increased use of this powerful and flexible statistical method.

Keywords

Longitudinal data, medical application, latent curve models

Disclosure: The research described herein was fully supported by a research grant (GRF 450607) from the Research Grants Council of the Hong Kong Special Administration Region, a grant (grant no. 931012) from the Health and Health Services Research Fund in Hong Kong and a research grant (grant no. 16000-3126133) from the Hundred Talent Program of Sun Yat-sen University. The authors have no conflicts of interest to declare.

Received: 21 January 2009 **Accepted:** 7 October 2009

Correspondence: Timothy Kwok, Professor, Department of Medicine and Therapeutics, Prince of Wales Hospital, the Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. E: tkwok@cuhk.edu.hk

Longitudinal data comprising repeated measurements of the same individuals on a number of occasions arise frequently in a wide range of fields: medicine, public health, psychology, biology and more. The main objectives of a longitudinal study are to characterise changes in the response of interest over time and to examine the selected co-variables that contribute to those changes. Traditional methods used to analyse longitudinal data are varied. Examples of these methods include autoregressive models, repeated measures multivariate analysis of variance, mixed-effects models, multiple regressions and so on.

Recently, there has been growing interest in models that have the ability to incorporate information concerning not only the group or population, but also changes in the individual. Latent (growth) curve modelling allows for the testing of complex models regarding developmental trends at both inter- and intra-individual levels. It has received increasing attention in medical research recently and has been well recognised as a useful longitudinal technique in the analysis of patterns of change.¹⁻⁶

Latent Curve Model

The latent curve model (LCM) is a method to model individual change, assess the effects of co-variables and assess the relationship among multiple outcomes, and take the measurement error into account. It provides a means of modelling developmental processes from both the inter- and intra-individual perspective. Generally speaking, LCM consists of two stages in modelling the patterns of change. In the first stage, the repeated measures of each

individual across time are fitted through a regression-type curve, which is either linear or non-linear. In the second stage, the focus of the analysis is on the latent growth factors, which are used to identify the individual's growth curve. The interest is no longer specifically on the original repeated measures observed, but on the unobserved latent growth factors that lead to the repeated measures. The LCM not only takes into account the mean of latent growth factors, which represent the group-level change, but also considers the variances that measure the degree of individual differences. This combination of group- and individual-level analyses is synthesised in the LCM procedure.

The interpretation of the model parameters in the LCM is illustrated with a simple two-factor linear model. The first-stage regression type equation, also called trajectory equation, is:

$$y_{it} = \alpha_i + t_i\beta_i + \varepsilon_{it} \quad (1)$$

where y_{it} is the measurement for individual i at time t , and α_i and β_i are, respectively, the random intercept and slope for individual i , t_i represents the sequential value of each time record and ε_{it} is the measurement error for individual i at time t .

In the second stage, the random intercept and slope are of interest, leading to the following two equations, called the intercept equation and slope equation, respectively:

and $\alpha_i = \mu_\alpha + \delta_{\alpha i}$ (2)

$$\beta_i = \mu_\beta + \delta_{\beta i}$$
 (3)

where μ_α and μ_β are the mean intercept and mean slope averaged across all individuals over time, $\delta_{\alpha i}$ is a residual reflecting the difference between the mean intercept μ_α and the individual intercept α_i and the individual slope β_i . The variances of $\delta_{\alpha i}$ and $\delta_{\beta i}$ measure the degree of diversity in individual specific intercepts and slopes from the mean intercept and slope.

A lot of extensions of the above basic LCM have been developed to meet real-world needs. For example, substantial co-variables can be included to account for their effects on latent growth factors.⁷ Non-linear LCMs can be used to capture the higher-order polynomial trajectories.^{2,7,8} For hierarchically longitudinal data, a multilevel LCM can characterise dynamic change of nested data,⁹ and for long-term heterogeneous data a mixture of LCMs can explore and identify different dynamic change patterns in different subject clusters.^{4,10}

Recent Applications of Latent Curve Models in Medical Research

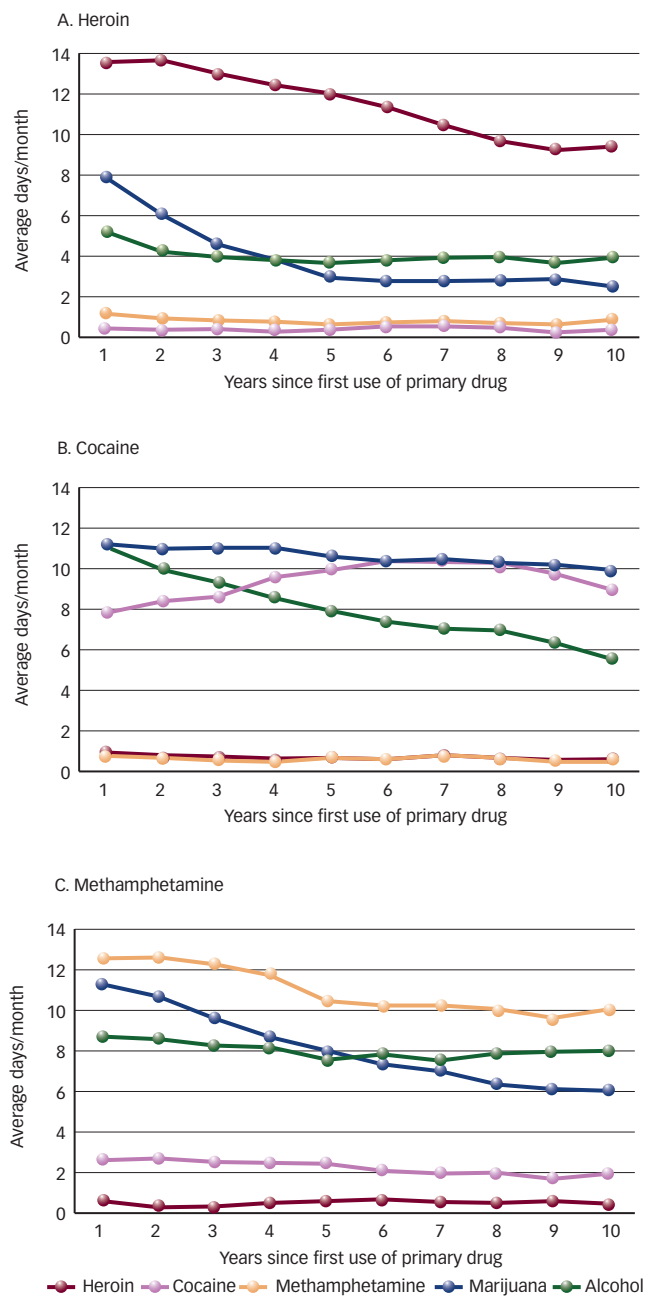
Linear Latent Curve Models

The linear LCM defined in equations 1–3 has been widely applied in medical research. The first illustrative example is the application of LCM in polydrug use. Brecht et al.¹¹ conducted an analysis to examine 10-year patterns of heroin, cocaine, methamphetamine, marijuana and alcohol use for primary users of heroin, cocaine and methamphetamine. Their main purpose was to investigate whether the trajectories of primary substance use over time are related to trajectories of other substance use. Non-overlapping samples from five studies that collected longitudinal information using the Natural History Instrument (NHI) were pooled in this analysis. Only those subjects who reported a primary drug problem of heroin, cocaine or methamphetamine were selected from each study.

Figure 1 depicts the trajectories of average number of days per month with use of each of the five substances for each of the primary drug sub-samples: heroin, cocaine and methamphetamine. It showed that the usage of non-primary heroin, cocaine or methamphetamine was very low during periods concurrent with use of these primary drugs. For users of all three primary drugs, however, there was consistently moderate usage of marijuana and alcohol. For each of the three primary drug sub-samples, a linear LCM was used to explore the patterns and relationships across drugs over time. The main finding was that the primary drug levels declined for heroin and methamphetamine users and showed relatively consistency over 10 years for cocaine users, while levels of non-primary drugs remained at consistently low levels or declined in tandem with the primary drug. This finding is consistent with the previous observed patterns in Figure 1.

Recently, Pan et al.³ applied the above LCM to a study of quality of life (QoL). It was a prospective cohort study of first disabling stroke patients conducted at the Prince of Wales Hospital (PWH) in Hong Kong. The study aimed to assess longitudinal behaviours of health-related QoL (HRQoL) in stroke survivors in relation to the changes in activities of daily living (ADL), handicap and post-stroke

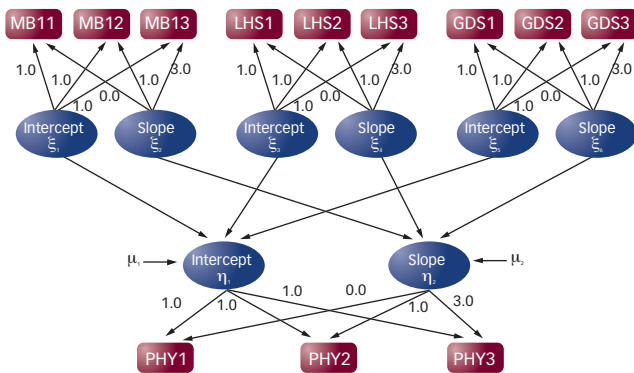
Figure 1: Observed Average Number of Days per Month of Use of Five Substances by Primary Drug Subsample



Source: Brecht et al., 2008.

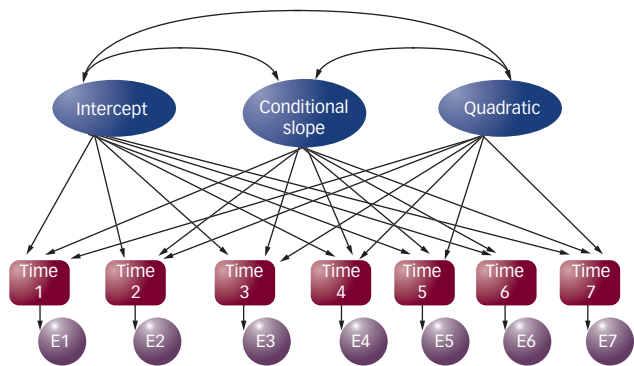
depression over time in the sub-acute phase of stroke recovery. Subjects were interviewed at three, six and 12 months after stroke to measure their levels of ADL, handicap, depression and HRQoL using the modified Barthel Index (MBI),¹² the London Handicap Scale (LHS),¹³ the Geriatric Depression Scale (GDS)¹⁴ and the World Health Organization (WHO) QoL questionnaire (abbreviated Hong Kong version).¹⁵ A linear LCM was proposed to examine how the dynamic changes in ADL, handicap and depression influenced dynamic changes in the four domains of HRQoL: physical health, psychological health, social interaction and environment. The LCM in analysing the physical health domain of HRQoL is depicted in Figure 2. The main finding of this analysis was that changes in mood in the sub-acute phase of the stroke recovery had the most significant effect on the

Figure 2: Latent Curve Model for Quality of Life of Stroke Patients



Source: Pan et al., 2008.

Figure 3: Quadratic Latent Curve Model of Depression Across Seven Time-points



Source: Beal et al., 2007.

HRQoL of the stroke survivors, while changes in basic functional status and handicap had less significant effects. In this longitudinal study, LCMs played an important role in delineating the relative contribution of basic functional status, handicap and depression in the HRQoL of stroke patients in a prospective manner.

Since the current care services in the sub-acute phase of stroke recovery do not specifically address the problem of post-stroke depression, this study suggests that, apart from drug treatment, there is a need to devise alternative strategies to manage post-stroke depression for better QoL for stroke survivors.

Non-linear Latent Curve Models

Basic LCMs can be extended to incorporate non-linear trajectories and effects of co-variates. With this extension, a higher-order polynomial is used to describe a non-linear pattern of dynamic change in individual characteristics. A longitudinal study about depression in multiple sclerosis (MS)⁷ is used for illustration.

The study goals focused on three research issues: to reveal the patterns of change in depressive symptoms over time; to identify substantial effects of co-variates such as age, type of MS (TMS), years since diagnosis of MS (YMS) and functional limitation (FL) on the trajectory of depression over time; and to examine the correlations between characteristics of change in FL and

depressive symptoms over the seven-year time period. The data were collected from 607 MS patients over a seven-year period, with initial recruitment in 1999 as part of an ongoing longitudinal study of QoL¹⁶ in chronic illness.

A non-linear trajectory in depression for the sample was suggested by an examination of randomly selected empirical growth plots. A three-factor LCM was used to model the trajectories of depression for the MS sample:

$$Y_{it} = \alpha_i + t_i\beta_i + t_i^2\gamma_i + \epsilon_{it} \quad (4)$$

where α_i , β_i and γ_i represent the random intercept, slope and quadratic slope for individual i , respectively. As depicted in Figure 3, the change pattern and the correlations between the characteristics of change in depressive symptoms were examined. Furthermore, predictors of change of depression were examined by regressing the intercept, slope and quadratic slope on the co-variates of interest, which leads to the following equations:

$$\alpha_i = \mu_\alpha + b_{\alpha 1}(\text{Age}) + b_{\alpha 2}(\text{TMS}) + b_{\alpha 3}(\text{YMS}) + b_{\alpha 4}(\text{FL}) + \delta_{\alpha i} \quad (5)$$

$$\beta_i = \mu_\beta + b_{\beta 1}(\text{Age}) + b_{\beta 2}(\text{TMS}) + b_{\beta 3}(\text{YMS}) + b_{\beta 4}(\text{FL}) + \delta_{\beta i} \quad (6)$$

$$\gamma_i = \mu_\gamma + b_{\gamma 1}(\text{Age}) + b_{\gamma 2}(\text{TMS}) + b_{\gamma 3}(\text{YMS}) + b_{\gamma 4}(\text{FL}) + \delta_{\gamma i} \quad (7)$$

Findings associated with the aforementioned three specific issues were obtained.

First, there was no significant increasing or decreasing trend in depressive symptoms, although it fluctuated over time for individuals. Second, younger age, longer time since diagnosis of MS, progressive forms of MS and greater extent of FL would result in greater depressive symptoms at time one. Third, FL showed an association with depression at all time periods, but other co-variates did not. In addition, gender did not predict the changes in depressive symptoms. These results indicated that screening for depression in all patients with MS was necessary and important.

Multilevel Latent Curve Models

In some circumstances, clinical trials involve a multilevel design, leading to hierarchically longitudinal data. For example, to evaluate the effects of a neighbourhood walking programme on QoL among older adults, Fisher and Li⁹ used a multilevel sampling scheme to collect a sample of neighbourhoods from a large metropolitan city, from which older adult residents were randomly recruited. This two-level design resulted in a nested data structure in which participants were clustered within neighbourhoods.

The substantive interest of this study focused on whether a six-month neighbourhood walking programme would improve neighbourhood-level QoL for senior residents. A two-level LCM of QoL, with individual- and neighbourhood-level data structures, is shown in Figure 4. All of the involved measures were assessed at baseline and at three and six months of the study period.

Compared with the control neighbourhoods, results from the two-level LCM indicated that physical, mental and satisfaction with life aspects of QoL were significantly improved over the course of the six-month intervention. The study concluded that it was feasible and

beneficial to implement a neighbourhood-based walking programme of low to moderate intensity in order to promote QoL among senior residents at a community level.

Mixture Latent Curve Models

Heterogeneity is commonly encountered in longitudinal analysis of medical research. For heterogeneous data there exist some latent classes under which the interested characteristics might present completely different change patterns.

A mixture of LCMs can be used to characterise the heterogeneity and to reveal specific change patterns for each distinctive latent class. Compared with the basic LCM, the additional tasks in applying mixture LCMs are to identify the number of latent classes, detect the membership of each individual observation and predict the probability of each individual falling in a specific class. To formulate the probability of an individual belonging to the latent class, the following multinomial logistic regression model was introduced. For $k=1,2 \dots K$,

$$P(C_i = k | x_{i1}, x_{i2}, \dots, x_{ip}) = \frac{\exp(a_{0k} + a_{1k}x_{i1} + a_{2k}x_{i2} + \dots + a_{pk}x_{ip})}{\sum_{j=1}^k \exp(a_{0j} + a_{1j}x_{i1} + a_{2j}x_{i2} + \dots + a_{pj}x_{ip})} \quad (8)$$

where K is the number of latent classes, C_i is the class membership for individual i , $x_{i1}, x_{i2}, \dots, x_{ip}$, are co-variables that may potentially influence the chance of individual i belonging to latent class k and $a_{0k}, a_{1k}, \dots, a_{pk}$ are corresponding regression co-efficients that reflect the importance of potential co-variables.

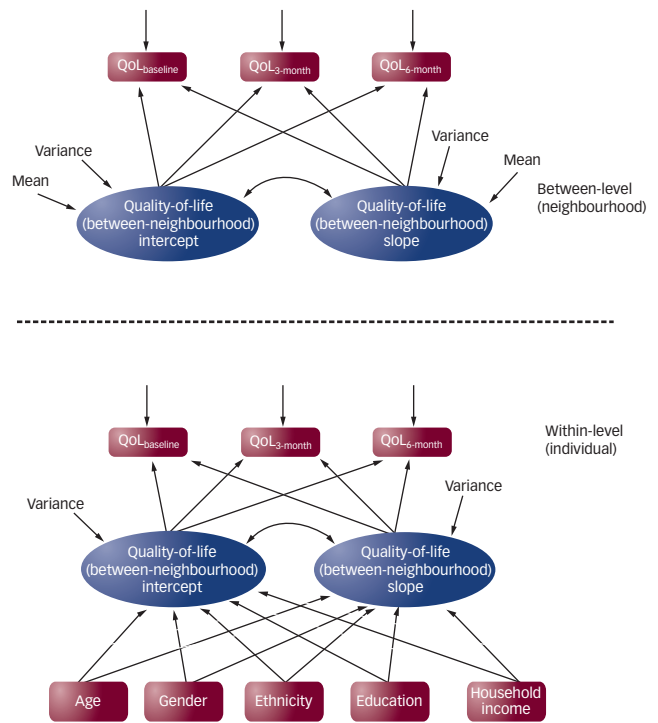
Hser et al.¹⁰ applied this model to examine long-term trajectories of drug use for primary heroin, cocaine and methamphetamine users. The data included 629 primary heroin users, 694 cocaine users and 474 methamphetamine users. The main outcome measure was the number of days using the primary drug per month.

As shown in *Figure 5*, the analysis of mixture LCMs revealed five distinct groups with different drug use trajectories over a 10-year follow-up: consistently high use, increasing use, decreasing use, moderate use and low use. In addition, primary drug type was significantly associated with different trajectory patterns.

Heroin users were most likely to be in the consistently high use group and cocaine and methamphetamine users were most likely to be in the moderate use group. The study also revealed that users in the high use group had earlier onset of drug use and crime, longer incarceration durations and fewer employed periods than those in other groups. Compared with other existing studies of drug addiction, the use of mixture LCMs in this analysis emphasised the heterogeneity of drug use patterns and the importance of understanding and addressing the full spectrum of drug use patterns over time.

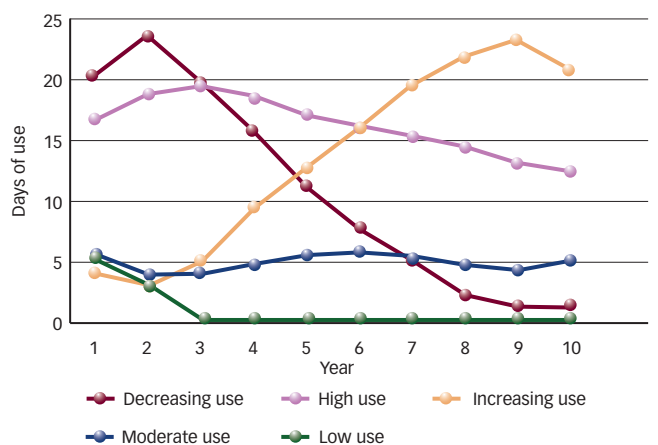
Another application of mixture LCMs is the analysis of depression in persons with myocardial infarction (MI). Elliott et al.⁴ analysed affect and event data from subjects post-MI in order to understand how mood and reactivity to negative events over time relate to diagnostic-level depression. In this study, 35 patients who had experienced an MI within the past year and were in treatment were investigated. The affect scores and event indicators (indicating

Figure 4: Multilevel Latent Curve Model for Quality of Life in Older Adults



Source: Fisher and Li, 2004.

Figure 5: Five-group Classification of Drug Use Trajectories Based on Mixture Latent Curve Model



Source: Hser et al., 2008.

presence of positive, negative and neutral events) of the patients were collected for up to 35 consecutive days.

The analysis of mixture LCMs suggested a two-class model for the MI patients: an 'optimist' class, with stable positive affect and declining perceived negative events, and a 'pessimist' class, with declining positive affect and continuing perceived negative events. Depressed subjects had a 92% chance of belonging to the 'pessimist' class compared with 62% among non-depressed subjects. This finding uncovered some hitherto unobserved structure in the positive affect and negative event data in this sample.

The key advantage of using mixture LCMs in this study was that persons who are most at risk of developing major or minor depression could be identified, which will assist in developing more specific interventions or treatments.

Conclusion

The LCM is an integrated and flexible approach that can be used to identify an appropriate growth curve for describing the developmental trend accurately and parsimoniously and to study the individual differences in changes. It has become a common and useful tool for the analysis of longitudinal data in behavioural, educational, social and psychological sciences. However, in medical applications LCMs have not been as popular as in other fields. There has been a lag between the recent development of statistical techniques and their widespread application in medical disciplines.

One reason why LCMs have been slow to move into the realm of medical sciences is that applied researchers are not familiar with this useful method and the associated user-friendly computer packages such as SPSS, SAS, Amos,¹⁹ EQS,²⁰ LISREL,²¹ Mplus,²¹ WinBUGS,²³ etc. It is believed that the medical field would benefit from an increasing awareness and use of these powerful and flexible methods. ■



Jun-Hao Pan is a Lecturer in the Department of Psychology at Sun Yat-sen University. His research interests are structural equation models, latent curve models, latent class models and statistical computing.



Xin-Yuan Song is an Associate Professor in the Department of Statistics at the Chinese University of Hong Kong. Her research interests are structural equation models, latent variable models, statistical diagnostics and semi-parametric and non-parametric methods.



Timothy Kwok is a Professor in the Department of Medicine and Therapeutics at the Chinese University of Hong Kong. His research interests are stroke rehabilitation, dementia, osteoporosis, nutrition and healthcare services for older people.

- Duncan TE, Duncan SC, Modelling the processes of development via latent variable growth curve methodology, *Struct Equ Modelling*, 1995;2:187–213.
- Bollen KA, Curran PJ, *Latent Curve Models – A Structural Equation Perspective*, New Jersey: John Wiley & Sons, Inc., 2006.
- Pan JH, Song XY, Lee SY, et al., Longitudinal analysis of quality of life for stroke survivors using latent curve models, *Stroke*, 2008;39:2795–2802.
- Elliott MR, Gallo JJ, Ten Have TR, et al., Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction, *Biostatistics*, 2006;6(1):119–43.
- Hser YI, Huang D, Chou CP, et al., Trajectories of heroin addiction: growth mixture modelling results based on a 33-year follow-up study, *Eval Rev*, 2007;31(6):548–63.
- Song XY, Lee SY, Hser YI, Bayesian analysis of multivariate latent curve models with nonlinear longitudinal latent effects, *Struct Equ Modelling*, 2009;16:245–66.
- Beal C, Stuifbergen AK, Sands DV, et al., Depression in multiple sclerosis: a longitudinal analysis, *Arch Psychiatr Nurs*, 2007;21(4):181–91.
- Meredith W, Tisak J, Latent curve analysis, *Psychometrika*, 1990;55(1):107–22.
- Fisher KJ, Li FZ, A community-based walking trial to improve neighborhood quality of life in older adults: a multilevel analysis, *Ann Behavioral Med*, 2004;28(3):186–94.
- Hser YI, Huang D, Brecht ML, et al., Contrasting trajectories of heroin, cocaine, and methamphetamine, *J Addictive Dis*, 2008;27(3):13–21.
- Brecht ML, Huang D, Evans E, et al., Polydrug use and implications for longitudinal research: ten-year trajectories for heroin, cocaine, and methamphetamine users, *Drug Alcohol Dependence*, 2008;96(3):193–201.
- Shah S, Vanclay F, Cooper B, Improving the sensitivity of the Barthel Index for stroke rehabilitation, *J Clin Epidemiol*, 1989;42:703–9.
- Harwood RH, Rogers A, Dickinson E, et al., Measuring handicap: the London Handicap Scale, a new outcome measure for chronic disease, *Qual Health Care*, 1994;3:11–16.
- Chiu HF, Lee HC, Wing YK, et al., Reliability, validity and structure of the Chinese geriatric depression scale in a Hong Kong context: a preliminary report, *Singapore Med J*, 1994;35:477–80.
- World Health Organization. Development of the WHOQOL-BREF from the WHOQOL-100. Geneva: WHO, 1997.
- Stuifbergen AK, Blozis SA, Harrison TC, et al., Exercise, functional limitations, and quality of life: A longitudinal study of persons with multiple sclerosis, *Arch Phys Med Rehabilitation*, 2006;87:935–43.
- Andresen EM, Malmgren JA, Carter WB, et al., Screening for depression in well older adults: Evaluation of a short form of the CES-D (center for epidemiological studies depression scale), *Am J Prev Med*, 1994;10(2):77–84.
- Kurtzke JF, A proposal for a uniform minimal record of disability in multiple sclerosis, *Acta Neurol Scand*, 1981;64(87):110–29.
- Arbuckle JL, Amos for Windows. Analysis of moment structures. Version 4.0, Small Waters, Chicago, 1999.
- Bentler PM, Wu EJC, EQS For Windows User's Guide. Multivariate Software. Inc, Encino, CA, 1995.
- Jöreskog KG, Sörbom D, LISREL for Windows. Lincolnwood, IL: Scientific Software International, 2006.
- Muthen LK, Muthen BO, Mplus – Statistical Analysis with Latent Variables User's Guide. Los Angeles, CA; 1998–2005.
- Spiegelhalter DJ, Thomas A, Best NG, et al., WinBUGS User Manual. Version 1.4. Cambridge, England: MRC Biostatistics Unit; 2003.