Application of Pattern-mixture latent trajectory modeling to assess longitudinal trajectory with non-ignorable missing data

Hiroko H. Dodge [1) 2)]
Changyu Shen [3)]
Mary Ganguli [4)]

1) Department of Public Health, Oregon State University, Corvallis, OR
2) Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA
3) Department of Medicine, Indiana University School of Medicine, Indianapolis, IN
4) University of Pittsburgh School of Medicine, Department of Psychiatry, Pittsburgh, PA

Corresponding author: Hiroko H. Dodge, 260 Waldo Hall, Dept. of Public Health, Oregon State University, Corvallis, OR 97331, USA. TEL/FAX: 541-342-3139. E-mail: Hiroko.Dodge@oregonstate.edu

EXTENDED ABSTRACT

Longitudinal designs, requiring follow-up of the same individuals over time, are increasingly common in epidemiological and demographical studies. However, missing data bias is a major problem in longitudinal studies where attrition is inevitable over time. For example, the frail are likely to miss or delay scheduled assessments for a variety of reasons, and a further problem is that the study outcomes themselves are often associated with the frailty such as dementia, disability, depression, and disease severity.

Restricting analyses to only the observed data could bias the results depending on the types of missingness. Little and Rubin (1) defined three types of missing-data mechanisms; 1. Missing Completely At Random (MCAR), which literally means missingness is completely random and does not depend on participants' characteristics, 2. Missing At Random (MAR), where missingness depends on participant's previously observed responses or observed characteristics, and 3. Missing Not At Random (MNAR), where missingness depends on unobserved outcome values (as well as possibly on observed values). Laird (2) and Little and

Rubin (1) also defined two general classes of missing-data mechanisms for likelihood-based approaches. A missing-data process is called ignorable if a likelihood-based approach provides valid inferences to the model parameters even when the missingness is ignored, while if not, called non-ignorable. Laird (2) showed that MAR is an ignorable missing-data mechanism. However, under MNAR, likelihood-based analyses that ignore the missing-data mechanism may be biased (non-ignorable missingness).

Two general classes of model based approaches were proposed to cope with non-ignorable missing-data: selection models and pattern-mixture models (3). The two models differ in the way to partition the joint distribution of outcome and missing indicator and both models have shortcomings; Since missing values, by definition, cannot be observed, the conditional distribution of missing indicator given outcome in selection models need to be based on an assumption which could still hold bias if this missing-data process is mis-specified. Pattern mixture models do not require specification of missing-data process and the marginal distribution of the response can be obtained by a weighted sum of the distribution within each pattern. However, one major problem in pattern mixture modeling approach is un-identifiability, or non-estimable parameters. For example, for a particular subgroup of the sample, if we observe only baseline data and there is no follow-up data at all, we cannot estimate the slope of the trajectory for that group.

Two major strategies to deal with the un-identifiability of pattern-mixture models are identifying restrictions (3-5) and model simplification (4). The first strategy assumes that the missing variable distribution (3) is equal to a function of corresponding identifiable distribution of some other patterns. The second strategy allows different patterns to share certain parameters so that the incomplete patterns can borrow information from patterns with more data points.

In this paper, we offer a practical solution to the un-identifiability problem by using latent trajectory analysis in the framework of pattern-mixture models. We applied this approach to estimate the longitudinal trajectories of a cognitive test, which taps memory function of individuals, over 12 years of follow-up. The approach presented here can be viewed as the model simplification described previously. Latent trajectory analysis (6-8) presented here account for the uncertainty involved in latent group assignment, which was originally proposed by Clogg (9).

Data

The Monongahela Valley Independent Elders Survey (MoVIES project) was a prospective epidemiological study of dementia from 1987 to 2002, set in the mid-Monongahela valley of southwestern Pennsylvania. The study background and methods have been reported previously in greater detail (10-12). Briefly the sample was selected by means of a 1:13 age-stratified (65-74, 75+), random sample of elderly individuals in 1987, identified through the voter registration lists. Entry criteria included age 65 years or older, being community-dwelling at the time of study entry, fluency in English, and at least a sixth-grade education. The last two conditions were designed to enhance interpretability of the neuropsychological tests. After giving informed consent, participants were interviewed by trained research associates. Study procedures were approved annually by the University of Pittsburgh Institutional Review Board. 1422 randomly selected from the voter registration list were assessed at study entry (Wave 1, 1987- 1989). At approximately two-year intervals thereafter, subjects were re-evaluated in a series of data collection waves. The assessment included cognitive testing with a battery designed to tap a range of cognitive domains affected by dementia. Here we focus on a memory test, the Word List Delayed Recall (WLDR) (10-item version developed for the Consortium to

Establish a Registry for Alzheimer's Disease (CERAD)(13). We estimated the mean trajectory on WLDR from Wave 1 (baseline) through Wave 6, over 12 years of follow-up, for the entire cohort, and also separately for age groups 65-75 and 75+ years, adjusting for non-ignorable missing-data bias.

At the upcoming PAA meeting, we will introduce the latent trajectory analysis and present a practical solution to the un-identifiability problem and how this method can be applied to various longitudinal data with non-ignorable missingness.

REFERENCE

1. Little RJ, Rubin DB. Statistical Analysis with Missing Data. New York: Wiley; 1987.
2. Laird NM. Missing data in longitudinal studies. *Stat Med* 1988;7:305-315.
3. Little RJ. Pattern-Mixture Models for multivariate incomplete data. *J Am Stat Assoc* 1993; 88:125-134.
4. Thijs H, Molenberghs G, Michiels B, Verbeke G, Gurran D. Strategies to fit pattern-mixture models. *Biostatistics* 2002;3:245-265.
5. Kennward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. *Biometrika* 2003;90:53-71.
6. Nagin DS. Analyzing developmental trajectories: a semi-parametric, group-based approach. *Psychol Methods* 1999; 4: 139-177.
7. Roeder K, Lynch KG, Nagin DS. Modeling uncertainty in latent class membership: a case study in criminology. *J Am Stat Assoc* 1999;94:766-776.
8. Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociol Methods Res* 2001;29: 374-393.
9. Clogg CC. "Latent class models". In Handbook of Statistical Modeling for the Social and Behavioral Science, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. 1995. New York: Plenum Press.
10. Ganguli M, Belle S, Ratcliff G, Seaberg E, Huff FJ, von der Porten K, Kuller LH. Sensitivity and specificity for dementia of population-based criteria for cognitive impairment: the MoVIES Project. *J Gerontol Med Sci* 1993;48: M152-M161.
11. Ganguli M, Dodge HH, Chen P, Belle S, DeKosky ST. Ten-year incidence of dementia in a rural elderly US community population: the MoVIES project. *Neurology* 2000;54:1109-1116.
12. Dodge HH**,** Shen C, Pandav R, DeKosky ST, Ganguli M. Functional Transitions and Active Life Expectancy Associated with Alzheimer's Disease. *Arch Neurol* 2003;60:253-259.
13. Morris JC, Heyman A, Mohs RC, Hughes JP, Belle Gv, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I: Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159-1165.