Commentary: Trajectories, Selection and Cumulative Causation

Andrew Pickles

Recent advances in life-course epidemiology have been considerable. Researchers now have a taste for examining a range of questions that involve contrasting combinations of patterns of exposure and patterns of impact on health outcomes that may be distant or occur over an extended time period, and data sets are becoming available where such questions can potentially be feasibly addressed. However, the statistical frameworks in which such ideas can be formalized and tested are currently limited in scope and the analytical problem has many different facets that require solution. Kaplan et al. present an interesting and surprisingly simple method of analysis that may offer useful insight. It suffers, however, from a variety of shortcomings.

One task in dealing with complex repeated measures is to simplify the outcome profile, and the reduction to a small number of trajectory classes offers an attractive option. Kaplan et al. estimate probabilities for each possible raw data trajectory under a simple model involving time updated exposures, and then sum these over a priori-defined classes of ‘similar’ trajectories. A contrasting model based approach, in common use in criminology and psychology, is to empirically identify trajectories. A contrasting model based approach, in common use in criminology and psychology, is to empirically identify trajectories. However, Kaplan’s method simply makes explicit assumptions implicit in such simpler models, while the Kaplan approach makes properly use of time-updated exposures in the determination of trajectories in a surprisingly straight-forward manner. However, many of these assumptions made in latent transition modeling merely make explicit assumptions implicit in such simpler analysis. For example, in the Kaplan et al. study, an explicit assumption is required as to the independence of some of the measurement properties of self-rated health.

A second major issue is the simultaneous consideration of both the state of health when alive and, when it occurs, the event of death. Such a joint consideration has been commonly cast in a Quality of Life Years framework, but in the Kaplan et al. study, the much more rudimentary categorization of death as a worsening of health has the appeal of simplicity. The occurrence of death also brings censoring of future observations. Having a distinct binomial model for inter-wave death membership probabilities or maximum a posteriori class assignment is then examined as a second step. This is potentially inefficient, particularly in a context of high levels of sporadic missing data or attrition. Alternatively, covariates can be linked to class membership probabilities through a multinomial model that forms part of the cluster analysis. While this works fine for baseline measures, the inclusion of later risk exposures raises serious concerns of allowing future values of exposure to influence trajectory probabilities where part of the trajectory has occurred prior to exposure. While latent trajectory models can be elaborated to consider latent transitions, with time updated risk factors influencing transitions in an appropriate time-ordered fashion, such models are complex and need considerable care in their formulation and implementation, and to the uninitiated seem to involve many arbitrary assumptions. By contrast, the Kaplan approach makes proper use of time-updated exposures in the determination of trajectories in a surprisingly straight-forward manner.

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2007; 36:548–549

doi:10.1093/ije/dym083

E-mail: andrew.r.pickles@manchester.ac.uk

Bibliography

and restricting the multinomial analysis of health change to survivors makes strong assumptions as to the conditional independence of these alternatives that may be hard to justify.

Finally, one of the major problems of longitudinal data analysis is the initial conditions problem. Simply put, this implies that for many sampling schemes, selection into the sample and an individual’s state of health at that time can only rarely be considered as exogenous to the process under investigation, with the consequence that different sampling schemes can deliver different results. For simplicity Kaplan et al. have selected only those in good health at the first wave. This may imply that, perhaps particularly among those older sample recruits, participants have been selected for unusually good health, regardless of their previous exposures and the sample will over-represent unusually resilient subjects. Those most vulnerable to the risk factors, already in poor health, may already have been excluded from the analysis. How do we generalize findings from such a selected sample to the whole population?

In short, the novel method of analysis presented by Kaplan et al. would appear to achieve a number of practical goals required by lifecourse researchers. However, further work is required to clarify how the identified effects formally relate to the process of the development of chronic disease, and the extent of generalizability of the resulting estimates, not to mention assessing their causal status.

References

4 Nagin DF, Tremblay RE. Trajectories of boys’ physical aggression, opposition, and hyperactivity on the path to physically violent and non-violent juvenile delinquency. *Child Dev* 1999;70:1181–96.