STRengthening Analytical Thinking for Observational Studies (STRATOS):
A SHORT SUMMARY OF 'AN OVERVIEW AND RECENT DEVELOPMENTS IN THE ANALYSIS OF MULTISTATE PROCESSES'

Malka Gorfine, Department of Statistics and OR, Tel Aviv University, Israel

Richard J. Cook, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

The paper by Gorfine et al. [1] describes the foundations of multistate modeling and illustrates the utility of this framework for studying dynamic disease processes involving distinct states of health when the times of transitions and paths taken carry scientific meaning. States may be defined to represent stages of a disease (e.g. remission, recurrence), hospitalization, discharge, or death. The review is motivated by the need for statistical guidance in such settings, where disease histories are often more complex than they are with a single time-to-event outcome.

The dynamic aspects of multistate processes are governed by transition intensities defined for each pair of states. These intensities represent the instantaneous risk of moving from a currently occupied state to another, given the prior history of the process, and have conceptual links to the hazard function in survival analysis. The paper explains the various ways the process history may affect the transition intensities, thereby defining different classes of models. In simple Markov models the risk of a transition depends only on the current state occupied whereas in semi-Markov models it also depends how long the current state has been occupied. More general models can incorporate richer dependencies on the process history and are often warranted for chronic diseases that feature both trends in time since disease onset, as well as recurrent periods in which symptoms are exacerbated.

The paper explains how familiar methods and software for survival analysis can be easily adapted for the analysis of multistate processes under right censoring. The Nelson–Aalen estimator is used to estimate cumulative transition intensities and the Aalen–Johansen estimator is used to estimate transition probabilities and state-occupancy probabilities over time. These quantities allow investigators to answer clinically meaningful questions such as the probability of being recurrence-free at a given time, the probability of death after relapse, or the expected amount of time spent in a given health state. The authors emphasize that these descriptive quantities are often more interpretable than transition-specific intensities, especially for reporting findings in applied medical research.

Regression modeling for multistate processes is also discussed. Intensity-based regression models, often formulated through multiplicative specifications for transition intensities analogous to the form of the Cox proportional hazards model, allow covariates to affect different transitions in different ways. This is useful because a predictor may increase the risk of disease onset while having little or no effect on mortality after onset. The paper carefully derives the likelihood framework needed when covariates are time-dependent and when censoring is present. A key message is that valid inference hinges on

assumptions about the censoring and covariate processes, especially when these processes may themselves be informative. This connects multistate modeling to broader joint-modeling problems in which marker trajectories, censoring mechanisms, and disease evolution may all be interrelated.

Beyond intensity-based approaches, the paper highlights the descriptive appeal of methods where covariate effects are modeled directly on marginal features such as state occupancy; pseudo-values can play a role here. These methods target quantities such as state-occupancy probabilities or restricted mean sojourn times directly, rather than building them indirectly through a full system of transition intensities. This is especially appealing when the scientific question concerns overall probability of occupying a state rather than the local behavior of transition rates. The pseudo-value approach is presented as a promising bridge between multistate modeling and marginal inference, although the authors note that important theoretical and practical issues remain open.

Another important area covered is when processes are only under intermittent observation. In many settings the exact transition times are not observable and instead the state is only ascertainable at clinic visits or other assessment times. Methods for handling such settings typically require stronger assumptions such as Markov assumptions with time-homogeneous or piecewise-constant intensities; modified likelihoods can be obtained for such settings and software such as the R package msm can be used to estimate transition intensities under intermittent observation. The psoriatic arthritis example in the paper illustrates how even when the exact timing of disease progression is unknown, multistate models can still estimate long-term state occupancy and identify risk factors associated with faster progression. The review also surveys random-effects and frailty approaches for capturing unobserved heterogeneity and dependence. These models are especially relevant when latent subject-specific factors induce dependence among a subject's event times. The authors discuss both conditional and marginalized frailty formulations for illness-death models under Cox-type regression, as well as accelerated failure time alternatives with shared random effects. Their discussion makes clear that unobserved heterogeneity is not merely a technical detail, but a substantive issue that can alter interpretation, prediction, and model fit.

A practical strength of the paper is its attention to implementation. The authors summarize available software in R and Python. This reinforces one of the review's broader goals: to connect modern methodology with tools that applied researchers can use. The paper repeatedly stresses the close relationship between multistate likelihoods and standard survival-analysis machinery, which lowers the barrier to adoption.

The paper concludes with a balanced assessment of the strengths and limitations of multistate modeling. Multistate models provide a coherent language for representing complex disease pathways, handling competing outcomes, and producing clinically interpretable predictions. At the same time, careful consideration is warranted when defining states, establishing the state space, specifying time scales, and ensuring the dependence structures are accounted for. The effect of model misspecification can be appreciable, especially when important covariates are omitted or when history dependence is oversimplified. The paper identifies several directions for future work, including causal inference, model validation, more flexible dependence structures, and machine-learning methods adapted to multistate data.

Overall, the paper serves as both a conceptual overview and a methodological guide. Its main message is that multistate models offer a powerful framework for understanding processes that unfold over time involving multiple transient and possibly absorbing states. Since survival analysis may be viewed as a special kind of multistate analysis involving two states (alive and dead), it is natural that related software enables fitting of complex multistate models. For researchers studying chronic disease processes the paper and related references provide a strong practical foundation for choosing, fitting, and interpreting multistate models.

References

1. Gorfine, M., Cook, R.J., Andersen, P.K., Therneau, T.M., Joly, P., Putter, H., Perme, M.P. and Abrahamowicz, M., 2025. An overview and recent developments in the analysis of multistate processes. Statistics in Medicine.