

Methods for Development and Validation of Individual- Level Clinical Prediction Models for Multi-State Processes: A Methodological Scoping Review

Chantelle Cornett

`chantelle.cornett@manchester.ac.uk`

University of Manchester

Alexander Pate

University of Manchester

Victoria Palin

University of Manchester

Glen P. Martin

University of Manchester

Research Article

Keywords: prediction models, multi-state model, methodology review, survival analysis, process history, transition probability, state occupation probability

Posted Date: September 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10704118/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Methods for Development and Validation of Individual-Level Clinical Prediction Models for Multi-State Processes: A Methodological Scoping Review

Chantelle Cornett¹, Alexander Pate¹, Victoria Palin^{1,2}, Glen P. Martin¹

¹ Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

² Maternal and Fetal Health Research Centre, Division of Developmental Biology and Medicine, The University of Manchester, Manchester, UK

Corresponding author: Chantelle Cornett (chantelle.cornett@manchester.ac.uk)

Abstract

Background

Multi-state models (MSMs) describe transitions between health states over time, supporting clinical prediction across multiple, potentially competing events. This paper aims to summarise methods for developing MSMs for individual-level clinical prediction and provide a structured guide to available approaches.

Methods

PubMed, Web of Science, Scopus, and arXiv were systematically searched and reference snowballing was conducted. Eligible studies were English-language, peer-reviewed articles proposing or validating statistical or machine-learning approaches for multi-state time-to-event models with covariates outputting individual-level risk estimates. Purely applied studies were excluded as well as papers focused on multimodal data, panel/intermittent/coarsened data, spatial analyses, reviews or software without methodological innovation.

Results

105 studies were included. Of these, 101 focussed on methods for model development. Papers developing methods for Markov processes were most common, with 55 papers, with 18 and 14 for semi-Markov and non-Markov, respectively. 11 papers described methods for prediction model development that are process-agnostic, including the use of pseudo-values ($n=7$). Fewer studies addressed missing data, hierarchical data, automated covariate selection, and sample size. Only 4 studies examined methods for model validation.

Conclusion

Although substantial methodological progress has been made in model specification and fitting, important gaps remain elsewhere in the prediction pipeline. Further methodological development is needed for sample size determination, missing data handling, and validation methods that assess discrimination, calibration, and uncertainty across multiple states and prediction times. This review is the first that systematically reviews available methods for producing MSM CPMs, serving as a useful guide for researchers developing or validating these models.

Keywords: prediction models; multi-state model; methodology review; survival analysis; process history; transition probability; state occupation probability.

1 Introduction

Clinical prediction models (CPMs) are regression models or machine learning algorithms that are used to predict the risk of future health outcomes, or the presence of current health outcomes [1]. The models take information on the patient (e.g. age, body mass index, smoking status, biomarkers) and use this to provide predictions that can be used to support clinical decision making. Such predictions can be used in clinical practice to identify individuals at high/low-risk of a disease or event, an example of this is the QRISK tool [2] that is used to estimate a patient's 10-year risk of cardiovascular disease (CVD). Based on the UK health guidelines if an individual were to have a QRISK-estimated CVD risk of 10% or more, then they are eligible for statin treatment to reduce this risk [3].

Most clinical prediction models are designed to predict a single outcome event. However, in many clinical settings, decision-making depends on understanding a patient's risk of experiencing multiple outcomes over time, and sometimes the order in which those outcomes occur. [4] For example, in multimorbidity prevention, clinicians may wish to estimate an individual's risk of developing type 2 diabetes, then atrial fibrillation, and subsequently cardiovascular disease. Statistically, this can be viewed as a multi-state process, in which individuals move through a sequence of clinically relevant states over time. Multi-state models provide a natural framework for representing these pathways and, in principle, for generating predictions about future transitions, state occupation, and disease progression. Although MSMs are often used to describe disease processes at the population level, they can also be used for individual-level clinical prediction.

In the multi-state individual-level prediction setting, the main quantity of interest is the transition probability: the probability that a patient in a given state will move to another state within a specified time horizon or stay in their current state, conditional on their covariates at the prediction time. [5] For example, one might wish to predict the risk of developing atrial fibrillation within three years of a type 2 diabetes diagnosis. In some settings, interest may also lie in the expected time spent in a state over a given horizon. MSMs are particularly useful because such predictions can be updated as patients move through different states over time.

As with all prediction models, the production of a MSM CPM involves the broader pipeline of development, validation, impact assessment, and clinical implementation. Developing a MSM CPM will often involve: (1) defining the state structure and transitions to represent the clinical pathway; (2) specifying the prediction setting, the time scale (e.g. clock-forward or clock-reset), the prediction horizon, and the covariates available at t_0 (and whether covariates will be updated over time); (3) addressing missing data, where required; (4) choosing and fitting an appropriate modelling approach; and (5) obtaining individual-level predictions for the target estimands. In many approaches, step (4) involves fitting transition-specific models (e.g. separate Cox-type models, flexible parametric hazards, or discrete-time approaches), including decisions about how competing events and censoring are handled, and step (5) then involves translating these fitted transition components into individual-level predictions (e.g. via the Aalen-Johansen estimator numerical integration, or simulation [5]). However, these steps are not absolute and may differ depending on the modelling strategy used. For example, pseudo-value approaches target clinically relevant prediction quantities more directly at fixed time horizons and do not necessarily require specification of a full set of transition-specific hazard models. [6] For validation of a MSM CPM, consideration then turns to quantifying the

predictive performance of the resulting transition probabilities, particularly calibration and discrimination, in the target population where the model is intended for clinical use. A recent review by Gorfine and colleagues addressed core multi-state model types (intensity-based, marginal, frailty approaches) and identified key challenges such as intermittent observation and emerging machine-learning applications. [148] However, the review focused on population-level descriptive analyses rather than covariate-specific individual-level prediction. This methodological scoping review differs in systematically mapping the full development and validation pipeline specifically for multi-state prediction models. To our knowledge, no existing review has focused on summarising this literature specifically where the interest is in developing a MSM for individual-level predictions across the full clinical prediction model pipeline. Therefore, the aim of this scoping review [7] is to provide an overview of all existing literature surrounding methods used to develop and validate multi-state prediction models for individual-level prediction. Our goal is to provide researchers with a comprehensive summary of the existing approaches they could use to produce multi-state individual-level CPMs in practice. We structure the review into stages of model development and validation, and we outline key advantages and disadvantages of available methods to provide guidance on choosing appropriate methods for each stage. We also call attention to opportunities for future research. Readers wanting information regarding the mathematical detail on multi-state models may refer to Andersen and Ravn (2023)[8] or Putter et al. [5].

2 Methods

2.1 Registration

The review protocol [9] was prospectively registered and is available on the Open Science Framework. The search strategy was revised: the protocol's search string included machine learning terms alongside general multi-state modelling terms, whereas the final strategy used a broader, method-agnostic string to capture the full methodological landscape. The grouping of evidence was also amended: the protocol described a descriptive and thematic synthesis organised chronologically, whereas the final synthesis used a more granular hierarchical grouping of pre-development, development, and validation, with development further subdivided by process type. We conducted and reported this scoping review in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews reporting guidelines. [10]

2.2 Search Strategy

We designed our search strategy to identify peer-reviewed journal articles that described methods for the development or validation of modelling multi-state processes for time-to-event outcomes. The full search strategies for all databases are provided in the Supplementary Material. The PUBMED search strategy was as follows: (("multi-state" OR "multistate" OR "multi state") AND ("development" OR "validation" OR "method*")) NOT ("multimodal*" OR "spatio*" OR "spatial*" OR "load*"). Results were filtered to English-language articles only. No date limits were applied.

Search terms were developed as a team based on our prior knowledge of this methodological area and were tested against key references identified *a priori* (Supplementary Material). Search terms were entered into PUBMED, Web of Science, ArXiv, SCOPUS and data was extracted on the 4th October 2025. The snowballing technique was also used to gather additional relevant papers with the aid of the Research Rabbit website [7]. Duplicates were removed using automatic deduplication by the Rayyan literature review online tool [11].

2.3 Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they described methods intended to support the development or validation of a multi-state model that could be used for individual risk prediction. Only papers published in peer-reviewed journals and written in English were considered.

Studies were excluded if they: (i) relied on panel/intermittently observed/coarsened data (since such data are rare to develop clinical prediction models); (ii) focused on multimodal data sources; (iii) did not focus on developing MSMs that can condition on predictors/covariates (since such conditioning is needed for individual risk prediction); (iv) presented only methodological comparisons/ reviews, solely described software packages, or contained no methodological development relevant to predictive modelling of multi-state processes (since we aimed to focus on papers developing new methods); or (v) contained only methods for discrete-time data (since routinely collected health data, upon which most prediction models are developed, [12] are predominantly captured in continuous time, albeit sometimes with irregular observation or coarsening). These exclusions were intended to keep the review focussed on methods specific to prediction modelling studies, and relevant to researchers working with electronic health records and other routinely collected clinical datasets. In addition, the review did not consider spatial processes. Pilot searches showed that our search terms retrieved a high volume of spatially focused papers, so terms with spatio* or spatial* in the title were excluded using the NOT operator.

2.4 Selection Process

Prior to title and abstract screening, GPM and AP each reviewed 5% ($n = 739$) of papers and CC reviewed each subset independently with the results compared for consistency. The three reviewers had high agreement (only 26 papers had a conflicting inclusion decision (1.27%)), and as such all remaining papers were screened by CC. For resource and time reasons, the team took a pragmatic approach to full-text screening and started with comparing decisions for 40 papers. Agreement was high (95%), and after discussions over differing decisions, a further 40 papers were assessed with 100% agreement. The remaining full-text screening was done by CC.

2.5 Data Collection and Data Items

After screening, the following information was extracted from each journal article (data collection form can be found in the Supplementary Material): the aim of the method, overview of the method, advantages, limitations, possibilities for future work as determined by the authors, and publication year. Publication years were extracted to allow for changing opinions in the “best practice” for that part of the modelling process. During full-text screening if a paper had unclear information, a second person (GPM or AP) would review the information, and a decision would be made jointly based on interpretation.

Papers were then placed into three main groups based on whether they included methods for pre-development, development or validation. The pre-development focused papers were then further split by which part of the development process they belonged to (i) missing data handling; or (ii) sample size; Development focused papers were further split into (i) data-driven variable selection methods; or (ii) model specification and fitting. Model specification

and fitting was further split into Markov processes, Semi-Markov processes, non-Markov, assumption-free and pseudo-values. Validation papers were split into discrimination and calibration.

2.6 Software Implementation Guidance

Although not the focus of this review, some detail is provided on software implementations available in R. This focus reflects the landscape of multi-state model development: the majority of widely used software packages for MSM specification, fitting, and prediction are implemented in R. Software descriptions and implementation guidance are included at the end of each methodological subsection, not as part of a systematic software review, but rather to aid researchers in future model development.

2.7 Glossary

Throughout the text many key statistical concepts and terms that are either intrinsic to multi-state modelling (such as Markov assumptions, transition probabilities, and the Aalen-Johansen estimator), or are commonly employed in the modelling, estimation, validation, or interpretation of multi-state processes are used. In the interests of brevity of the main text, we provide a glossary table in the Supplementary Material (Supplementary Table 1).

3 Results

3.1 Study Selection

Initially, 18207 studies were identified from four databases: Web of Science (5674), PubMed (3942), SCOPUS (8071), and ArXiv (520). After removing 7161 duplicates (348 manually and 6813 by Rayyan), 11046 studies remained for screening (**Error! Reference source not found.**). Following the initial abstract and title screening, 308 studies were assessed for eligibility.

As listed in **Error! Reference source not found.**, a total of 216 studies were excluded for reasons related to eligibility criteria. This process resulted in 105 studies being included in the final review, with the additional 13 studies identified through forward and backward snowballing techniques. 55 addressed Markov processes, 18 semi-Markov, 14 non-Markov, 11 were process-agnostic, and 7 used pseudo-values (see Table 1 for a list of abbreviations and Supplementary Materials for Glossary of key terms).

Studies that may appear to meet the inclusion criteria, but were excluded include: “A Note on Variance Estimation of the Aalen–Johansen Estimator of the Cumulative Incidence Function in Competing Risks, with a View towards Left-Truncated Data” (2009), [13] which was rejected as it did not condition on predictors, and “Progressive multi-state models for informatively incomplete longitudinal data” (2011), [14] which focused on causal effects.

Table 1 List of abbreviations included in the review

Abbreviation	Definition
AIC	Akaike information criterion
AUC	Area under the curve
BLR-IPCW	Binary logistic regression with inverse probability of censoring weighting
BUGS	Bayesian inference using Gibbs sampling
CPM	Clinical prediction model
CVD	Cardiovascular disease
EHR	Electronic health record
EM	Expectation-maximisation
GEE	Generalised estimation equation
ICS	Informative cluster size
IPCW	Inverse probability of censoring weighting
JAGS	Just another Gibbs sampler
LASSO	Least absolute shrinkage and selection operator
MAR	Missing at random
MCMC	Markov chain Monte-Carlo
MI	Multiple imputation
MLR-IPCW	Multinomial logistic regression with inverse probability of censoring weighting
MNAR	Missing not at random
MSM	Multi-state model
NIHR	National Institute for Health and Care Research
NICE	National Institute for Health and Care Excellence
OSF	Open Science Framework
PLS	Partial least squares
PMM	Predictive mean matching
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews
QRISK	Cardiovascular disease risk prediction tool (UK)
ROC	Receiver operating characteristic
SPLS	Sparse partial least squares
STAR	Structured hazard multi-state regression

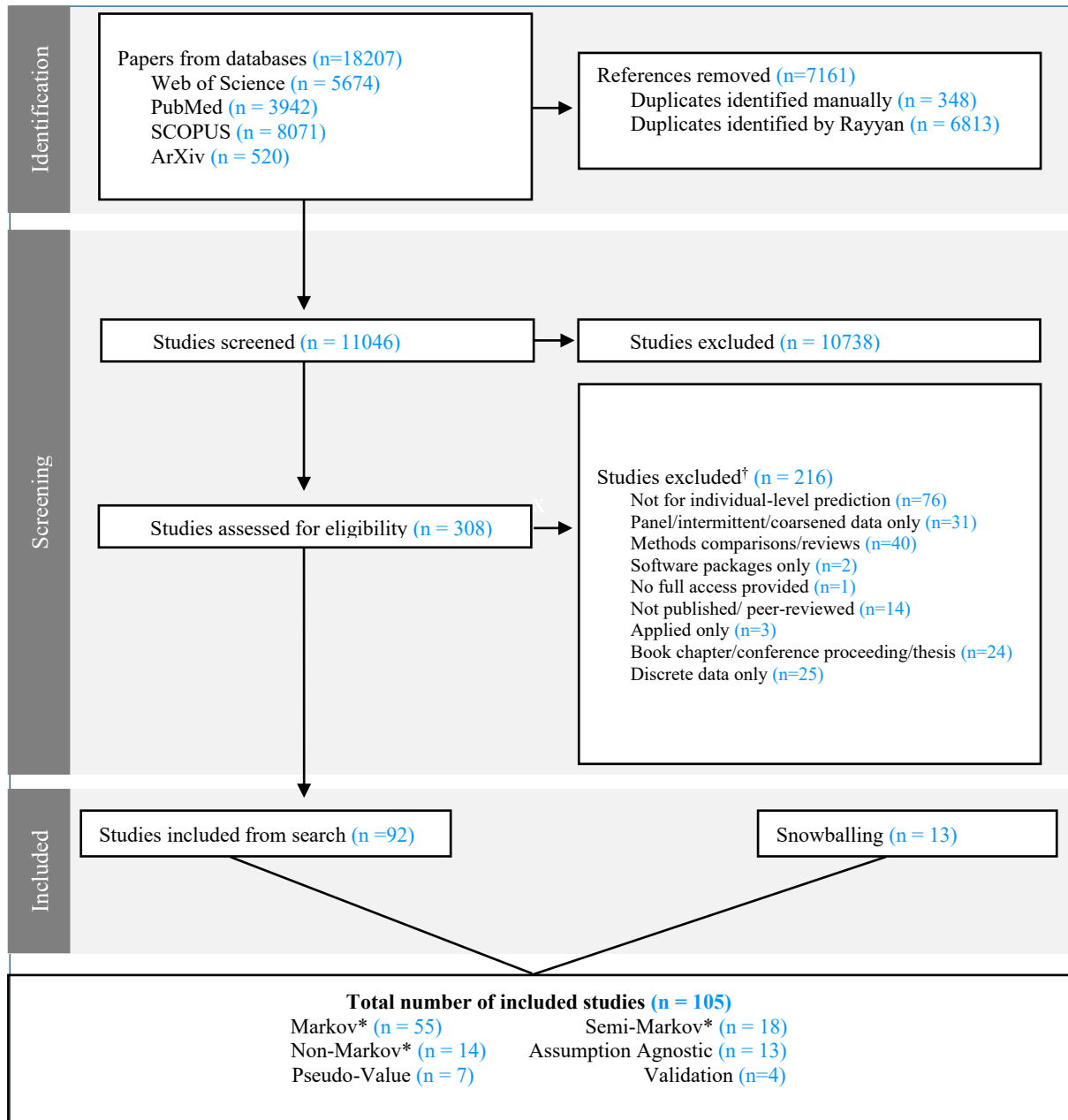


Figure 1 PRISMA flow diagram for methodological studies relating to multi-state processes. †Includes primary reason for exclusion only. *Papers may belong to more than one category.

3.2 Pre-Development

3.2.1 Sample Size

Determining the required sample size for a multi-state CPM is more complex than for single-outcome models: [4, 15] researchers must ensure adequate events for each transition, accommodate potentially sparse absorbing-state transitions, and consider the number of predictors per model. Despite this, none of the 105 development-focused papers in this review directly proposed sample size methods for multi-state CPMs. This gap is discussed further in Section 4.

3.2.2 Missing Data

Missing-data methods support valid estimation of MSM CPM prediction estimands by addressing missing covariates and uncertain transition times. While missing data are a common issue across prediction modelling, MSM CPMs pose additional challenges because

predictions depend not only on predictor values but also on the timing and ordering of transitions between states. Missingness may therefore affect both the covariate information used for prediction and the observed event-history process used to estimate transition risks. We identified two papers that focused on missing data handling for Markov multi-state models.

The first, [16] develops a likelihood-based approach for multi-state models with partially missing covariates. The authors construct the observed data likelihood by jointly modelling outcomes and covariates under missing completely at random/ missing at random (MAR), resulting in unbiased, efficient estimates and avoiding the convergence issues of complete-case analysis or naive imputation.

The other [17] compares multiple imputation (MI) strategies for missing event times, including predictive mean matching (PMM), subgroup-specific PMM, linear regression MI and a compatibility-preserving PMM. PMM, particularly subgroup-specific PMM, shows low bias and good precision under MAR and improves efficiency over complete-case analysis. The study found considerable bias under MNAR and found that MI using draws from a linear imputation model led to more bias than PMM when estimating transition intensity model parameters.

3.3 Model Development

3.3.1 *Data-Driven Variable Selection Methods*

Data-driven variable selection methods aim to reduce overfitting and improve estimation across multiple transition-specific models, particularly when many predictors are considered or some transitions are infrequent. Table 2 provides a summary for each paper included in this section.

Table 2 Summary of literature for Markov algorithmic covariate selection

Title	Year	Aim	Method	Benefits	Limitations
Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models [18]	2008	To introduce reduced-rank regression techniques in multi-state models to address the difficulty of estimating many covariate effects, particularly for event-poor transitions	Applied reduced-rank Cox regression for transition intensities, borrowed information from event-rich transitions, and used resampling methods to estimate standard errors and prediction probabilities	Reduced the number of parameters, improved stability of estimates, yielding interpretable prognostic scores, and enabled practical analysis of complex settings	Computationally intensive due to simulation and resampling, standard error formulas not available for semi-Markov models, prediction probability inference relies heavily on simulation
Structured fusion lasso penalized multi-state models [19]	2016	To introduce a structured fusion Lasso penalised framework for multi-state models to enable cross-transition effect analysis and variable selection in a data-driven way	Combined L1 and fusion penalties within partial likelihood and piecewise constant baseline hazard models, estimated via penalised iterative reweighted least squares with local quadratic approximation, selected tuning parameters via grid search using AIC	Identified structural relationships between transitions, unifying variable selection and cross-transition effect estimation	Current approach limited to linear effects; extension to nonlinear (e.g., penalised splines) would be complex due to many penalty parameters
Risk Factor Identification in Heterogeneous Disease Progression with L1-Regularized Multi-state Models [20]	2021	To develop an L1-regularised multi-state model framework for simultaneous parameter estimation and variable selection	Applied L1-regularised partial likelihood with a one-step coordinate descent algorithm using local quadratic approximation, specified Cox models for each transition	Improves computational efficiency, identifies significant risk factors more accurately than existing approaches, enhances prediction of transition probabilities, and performs well even in small-sample settings	Current approach assumes constant coefficients over time, cannot jointly handle categorical variables without group penalties, and does not yet address time-dependent or semi-Markov extensions
Multistate analysis of multitype recurrent event and failure time data with event feedbacks in biomarkers [21]	2022	To propose a flexible and computationally efficient modelling framework for analysing multitype recurrent events alongside terminal events	Extended Cox regression with both constant and time-dependent coefficients (estimated via regression splines), allowed event-type and event-order dependence, modelled event feedback, avoided distributional assumptions for random effects, and developed asymptotically unbiased estimating equations with derived variance estimators	Handled complex event feedback structures, supporting hypothesis testing and confidence interval construction without bootstrapping, improved computational efficiency over local polynomials, and offered flexibility for recurrent/terminal event processes	Current framework assumes Markov models, but feedback mechanisms may violate this assumption; extensions to semi-Markov or non-Markov models and formal testing for feedback effects are needed
Penalized Estimation of Frailty-Based Illness-Death Models for Semi-Competing Risks [22]	2024	To develop penalised frailty-based illness-death models to better capture semi-competing risks and improve interpretability and prediction in clinical application	Applied fusion regularisation to regression parameters across hazards, allow penalization of baseline hazards, compared frailty and non-frailty models, and established theoretical statistical properties of the penalised estimator	Improved model fit and parsimony, enabled sharing of covariate effects across hazards, supporting individualised risk prediction, and is adaptable to high-dimensional data sources	Current work assumes parametric baseline hazards and gamma frailty; extensions to spline-based baselines, alternative frailty distributions, and post-selection inference remain future directions

Markov multi-state models have been extended with algorithmic covariate selection and cross-transition regularisation to address instability arising from the need to estimate many transition-specific effects, particularly for event-sparse transitions [18-22]. One method uses reduced-rank proportional hazards regression for transition intensities to reduce dimensionality, allowing prediction in complex multi-state settings [18]. Other methods have also been used to select predictors and to incorporate shared structure among transitions. For example, including structured fusion-lasso multi-state models that combine sparsity and cross-transition fusion penalties within Cox-type partial likelihoods and L1-regularised multi-state models with coordinate-descent algorithms for variable selection [20]. Similar developments for Markov recurrent-event/terminal-event systems allow for flexible effect modelling (including spline-based time-varying coefficients) within a multi-state Cox framework via estimating equations [21] without the need for computationally demanding inference. More recently, penalisation has been applied to frailty-based Markov illness–death models for semi-competing risks, using fusion regularisation to share covariate information across hazards, thereby improving prediction [22].

For individual-level multi-state prediction models, reduced-rank proportional hazards regression [18] provides the most user-friendly method of variable selection, combining automatic regularisation with simulation-based trajectory prediction. L1 or fusion-based feature selection [19, 20] can precede this to identify relevant predictors. If time-varying covariate effects are scientifically important, the flexible spline-based approach complements this framework without requiring fundamental restructuring. [21] Penalisation, alongside frailty-based Markov illness–death models, adds complexity beyond standard multi-state prediction and should be adopted only when genuinely applicable. [22]

3.3.2 *Model Specification and Fitting*

3.3.2.1 Markov Models

Markov multi-state methods in the reviewed literature [23-49] were mainly concerned with how to specify the transition structure and fit the model to longitudinal event-history data so that clinically relevant predictions could be derived. For someone developing a multi-state prediction model, these methods are useful because they provide a framework for estimating not only whether an event occurs, but which transition occurs, when it occurs, and how long patients are expected to remain in different states. Across the literature, the main outputs included transition hazards, transition probabilities, state occupation probabilities, and expected time spent in particular states.

Many papers specified the multi-state process through transition intensities, using either parametric or semiparametric regression models. [23-26, 33-35, 48, 50-54] Parametric approaches assumed specific forms for the baseline transition hazard, such as exponential, Weibull, or piecewise-constant models, usually fitted by maximum likelihood [23, 26, 48]. For someone fitting a multi-state prediction model, these approaches are helpful because they provide a fully specified model from which transition probabilities and related prognostic summaries can be derived directly, often with relatively straightforward estimation. In contrast, when the modeller wants greater flexibility in the baseline transition process while still quantifying how predictors affect movement between states, then semiparametric approaches can be used, most commonly Cox-type proportional intensity models, which leave the baseline hazard unspecified while estimating transition-specific covariate effects [24, 25]. We identified a related strand of research focused on transition and state occupation

probabilities, using nonparametric product-integral methods such as the Aalen–Johansen estimator [24, 33–35]. These methods are particularly useful when the aim is to estimate absolute probabilities of being in a given state over time, which are often closer to the quantities clinicians want for prognosis than transition-specific hazard ratios alone.

The literature also included a wide range of extensions to support model fitting under various data conditions. [23–49] Some methods adapted standard Markov models to settings with incomplete cohort information or complex sampling designs, such as nested case-control studies [27, 31], thereby enabling the modeller to estimate transition effects even when full follow-up data are not available for all individuals. Hidden Markov approaches addressed misclassification, which is useful when the recorded state may not equal the true underlying disease state [28, 46, 49]. Frailty and related latent-effect models were used to represent unobserved heterogeneity across individuals or transitions [29, 47], helping the modeller account for dependence that is not captured by measured predictors alone. Further extensions allowed for time-dependent covariates, interval- or left-censoring, recurrent-terminal event structures, and joint modelling with longitudinal biomarkers [32, 37–44], which is particularly helpful when prognosis needs to be updated over time using evolving clinical information rather than baseline covariates only. Together, these methods show that model specification in this literature was not just about selecting a hazard form, but about deciding how best to represent incomplete observation, latent structure, and changing patient information within the fitted prediction model.

Additional work considered clustered or hierarchical data, where transition outcomes may be correlated within centres, hospitals, or other groupings. Approaches included mixture Markov regression models, frailty-based multi-state models, hierarchical Bayesian illness–death models, and nested frailty formulations [55–61]. For those fitting multi-state prediction models, these methods are useful because they allow predictions and uncertainty estimates to reflect shared centre-level or subgroup-level structure, rather than treating all patients as independent. This is especially relevant in multicentre datasets, where failing to account for clustering may lead to overconfident predictions or hide between-centre variation in transition dynamics (Table 3).

Table 3 Literature relevant to handling clustering in Markov MSMs

Title	Year	Aim	Method	Benefits	Limitations
Mixture multi-state Markov regression model [59]	2007	To develop a mixture multi-state Markov regression model that accounts for unobserved heterogeneity in disease progression through an initial mixture distribution and covariate-dependent transition rates	Used a continuous-time homogeneous Markov model with an unobserved mixture distribution for the initial state and exponential regression models for transition intensities, estimated via maximum likelihood	Reduces unobserved heterogeneity, improves fit over standard homogeneous models, and identifies covariate effects on specific transitions for more accurate clinical classification	Requires exponential assumptions, depends on correct specification of the initial mixture distribution
Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units [57]	2012	To develop multi-state frailty models to account for cluster-level heterogeneity in competing risks and multi-centre data	Specified semi-Markov and non-homogeneous Markov models with semiparametric or parametric estimation, provided model selection measures	Captures centre effects and unobserved heterogeneity, improves predictive accuracy compared to models without frailties, allows cluster-predictions, and provides practical software for implementation	Requires large datasets for frailty variance estimation, prediction updating with time-dependent covariates remains an open problem
Bayesian Semiparametric Analysis of Semicompeting Risks Data: Investigating Hospital Readmission After a Pancreatic Cancer Diagnosis [55]	2015	To develop a unified Bayesian framework for semi-competing risks data to estimate regression parameters, characterise within-subject dependence, and predict outcomes	Implemented Bayesian illness-death models (Markov and semi-Markov) with gamma frailties, MCMC estimation, and posterior predictive distributions	Provides a comprehensive Bayesian approach that incorporates prior information, quantifies uncertainty, enables prediction, and flexibly extends to complex structures	Interpretation depends on time scale and frailty assumptions, with non-terminal event distributions not identifiable without unverifiable assumptions
Hierarchical Models for Semicompeting Risks Data With Application to Quality of End-of-Life Care for Pancreatic Cancer [56]	2016	To develop hierarchical models for semi-competing risks	Used illness–death models with Bayesian hierarchical extensions, incorporating patient- and cluster-level random effects	Captures dependence between non-terminal and terminal events, accounts for clustering and heterogeneity	Computationally intensive, requires careful specification of priors, results sensitive to model assumptions
Correlated multistate models for multiple processes: an application to renal disease progression in systemic lupus erythematosus [58]	2018	To develop a joint modelling framework that links recurrent events, a terminal event, and an underlying latent multi-state process to better characterise individual event histories	Specified a latent continuous-time Markov process driving recurrent-event and terminal-event intensities, and estimated parameters using likelihood-based inference incorporating shared latent-state structure	Captures unobserved heterogeneity, allows dependence between recurrent and terminal events, and yields improved modelling of individual event trajectories compared with standard frailty or marginal approaches	Computationally intensive, relies on correct specification of the latent multi-state structure, becomes difficult to scale as the number of latent states or event types increases
Multistate models with nested frailty for lifetime analysis: Application to bone marrow	2024	To propose and evaluate multi-state frailty models to capture heterogeneity between groups and dependencies in clustered transition data	Specified shared frailty (MMSF2) and nested frailty (MMNF) models, and applied parametric and semiparametric estimation	Accounts for both group-level and transition-level heterogeneity, improves modelling of clustered survival data, and identifies transition-specific covariate effects	MMNF assumes equal variance across transition–cluster interactions, which may oversimplify heterogeneity, model complexity increases with number of states and transitions,

transplantation recovery patients [60]					and interpretation of frailty estimates requires caution
Bayesian path specific frailty models for multi-state survival data with applications [61]	2015	To develop a Bayesian gap-time multi-state frailty model with path-specific random effects that captures dependence between transitions and enables estimation of path and survival probabilities	Specified multinomial logistic models for immediate child states and gamma frailty piecewise-exponential gap-time models for each path, derived conditions for posterior propriety under improper priors, and implemented inference via a Gibbs sampler	Provides flexible non-proportional hazards via generalised odds-rate margins, yields path and survival probabilities as direct outputs of the Gibbs sampler, and performs well in simulations and real data	Assumes a unidirectional progressive path structure with shared frailty across transitions in a path, can struggle with posterior propriety and sparse paths without informative priors, and becomes computationally demanding as the number of states and paths increases

A smaller group of literature explored alternative specification and fitting approaches. [62–66] Flexible spline-based and computationally efficient continuous-time models aimed to improve fit while reducing restrictive functional form assumptions or computational burden [62, 63]. Machine-learning approaches, including neural-network methods and approaches combined with Aalen–Johansen-type estimation, were used to capture more complex linear and nonlinear relationships between predictors and transition behaviour [64, 65]. These methods may help someone fitting a multi-state prediction model when the goal is to improve predictive accuracy or calibration in the presence of predictor values for an individual that are extreme and that may be hard to capture with conventional transition models. Quantile-regression approaches offered a different type of prediction target by estimating conditional quantiles of transition or sojourn times [66]. This is helpful when the practical question is not just the probability of transition, but when that transition is likely to occur, for example, when estimating median time to progression or time spent in an intermediate state.

Across these methods, there remain some areas for methodological development. Although many approaches estimate transition hazards or probabilities successfully, there is a scarcity of methods that allow models to output prediction summaries that are most useful in practice, such as counterfactual prediction. Many of the most flexible approaches were computationally demanding and difficult to implement routinely, which may reduce their accessibility for applied researchers developing multi-state prediction models.

Overall, the reviewed literature shows that Markov multi-state models can be specified and fitted using a wide range of approaches, from standard parametric and semiparametric intensity models to nonparametric, latent-variable, hierarchical, and more flexible prediction-oriented methods. For researchers, these approaches provide tools for estimating predictions in more complex data settings. However, important challenges remain in aligning fitted models with clinically useful estimands and in balancing model flexibility with robust and interpretable estimation.

Software implementing model fitting approaches is available in R: the flexsurv package implements flexible parametric (including spline-based) transition hazards models.[67] Cox-type transition models can be fitted using the standard survival or coxph infrastructure within R. [68] Stan and BUGS/JAGS can be used for fully Bayesian Markov multi-state specifications. Additionally, for clustered or hierarchical data, the frailtypack package supports shared frailty models for multi-state settings with parametric and semi-parametric estimation. [69] The R package coxme allows mixed-effects Cox models that can be applied to transition-specific models. The JM and JMbays2 packages support joint models with frailty-like random-effects in R. [70]

3.3.2.2 Semi-Markov

Semi-Markov multi-state methods are used when future transition risk depends not only on the current state but also on the time already spent in that state. For someone fitting a multi-state prediction model, this is useful when prognosis is duration-dependent, since it allows prediction of transition probabilities, state occupation probabilities, and time-in-state summaries that vary with elapsed sojourn time [71, 72]. Supplementary Table 5 contains all literature found pertaining to semi-Markov methods [39, 45, 71–84].

The reviewed literature included semi-Markov regression models that incorporated covariates, censoring, and duration dependence through likelihood-based, semiparametric, or parametric waiting-time formulations [73–77]. These methods help model developers by

allowing future risk to depend on both patient characteristics and time since entering the current state, rather than assuming all patients in the same state have the same prognosis. A second group of papers focused on estimating prediction quantities and their uncertainty, including sojourn distributions, transition probability estimators, and resampling-based inference [76, 78]. This is important because semi-Markov prediction relies on derived quantities that need appropriate uncertainty estimation. A third group of papers extended semi-Markov modelling to more complex prediction settings, including transformation and relative survival models, hidden or informatively censored semi-Markov processes, Bayesian and shrinkage-based approaches, joint longitudinal models, and spline-based covariate effects [39, 45, 71, 72, 79-84]. These extensions broaden the range of data structures and predictor relationships that can be handled in duration-dependent prediction.

There are some weaknesses of semi-Markov modelling methods. First, semi-Markov prediction depends critically on accurate information about time in state, so censoring, delayed entry, and misclassification can directly distort estimated sojourn distributions and the predictions derived from them [73, 76, 78, 81, 85]. Second, increasing flexibility in duration-dependent modelling often creates a trade-off with stable, interpretable estimation [39, 71, 72, 77, 83, 84]. Third, some approaches remain difficult to fit routinely in large registry or EHR datasets because they require substantial sample sizes or computationally intensive estimation [18, 71, 72, 77].

Semi-Markov models are used when prediction depends on how long a patient has remained in a state, but robust estimation and efficient implementation are challenging in practice. Software implementations include the Semimarkov R package [86], which provides dedicated parametric model fitting. The flexsurv package can also implement semi-Markov models by fitting transition-specific parametric sojourn distributions [67], and simulation-based state occupation probabilities are available via the mstate package. [87] Bayesian semi-Markov models can be fit using Stan.

3.3.2.3 Non-Markov

Non-Markov approaches are most relevant when transition probabilities or state-occupation probabilities depend on patient history beyond the current state. For someone fitting a multi-state prediction model, this is useful when prognosis cannot be estimated adequately by the current state alone, particularly for dynamic prediction from intermediate states. Supplementary Table 6 contains all literature found pertaining to non-Markov methods. [88-102]

Most non-Markov methods were developed to estimate history-informed transition and state-occupation probabilities without specifying a full transition intensity model, often using landmarking, smoothing, or other nonparametric estimators [88-92, 94] [95-98].

Within this body of literature, we have identified some areas for further methodological development. Reducing bias due to Markov misspecification often comes at the cost of greater variance, making it difficult to assess when non-Markov methods offer an advantage over simpler Markov approaches [90, 91]. Additionally, although methods allow covariate-specific prediction, their ability to handle more complex covariate structures, time-varying information, and higher-dimensional settings is limited [94-97, 99].

Non-Markov methods provide useful tools when prediction depends on more than the current state alone, especially for dynamic prediction from intermediate states, but efficient and reliable estimation of history-dependent risks continues to be challenging. Software for

implementation includes the `mstate` package, which implements the landmark Aalen-Johansen estimator.[87] Landmarking for dynamic prediction is also supported by the `landpred` R package. [103]

3.3.2.4 Unspecified Structures

Existing methodology for analysing multi-state event-history data with unspecified structures includes both transition-specific hazard models and estimator-based approaches that directly target clinically interpretable prediction quantities. Supplementary Table 7 contains the literature found pertaining to unspecified structure methods. [104-114] For someone fitting a multi-state prediction model, these methods are useful because they provide ways to estimate transition probabilities, state occupation probabilities, time-in-state summaries, and dynamic predictions without relying on a single process-specific framework.

The reviewed literature included transition-specific models, multiple-spell approaches, nonparametric estimators of trajectories, semiparametric regression for sojourn times, inverse-probability weighted methods for left-truncated settings, landmarking approaches for dynamic prediction, multiple-time-scale models, and more flexible structured additive or boosting-based methods [104-112].

Many methods were developed for simple state structures, such as illness-death or semi-competing risks models, limiting their use in more complex pathways with multiple intermediate states or recurrent transitions [105, 106, 111-113]. There is also limited guidance on handling multiple time scales, even though choices about time since entry, illness duration, or calendar time can affect predictions [108, 113]. Finally, more flexible methods, such as multi-state boosting, aim to capture complex covariate effects, history dependence and misclassification, but are computationally demanding and can be difficult to diagnose for convergence and overfitting in realistic datasets [109].

3.3.2.5 The Pseudo-Value Approach

Pseudo-value methods provide an alternative approach to model specification and fitting in multi-state prediction by targeting fixed-horizon transition probabilities, state occupation probabilities or related cumulative quantities at fixed time horizons, rather than modelling transition hazards and then deriving probabilities under a particular process assumption. Table 4 contains all the literature found regarding pseudo-value approaches. [6, 95, 114-119] For someone fitting a multi-state prediction model, this is useful when the main aim is to predict a clinically interpretable absolute risk at a chosen time point, such as the probability of being in a particular state by time, without fully specifying the underlying transition process [6, 115].

Table 4 Summary of literature regarding pseudo-values

Title	Year	Aim/Purpose	Method	Benefits	Limitation
Generalised linear models for correlated pseudo-observations, with applications to multi-state model [120]	2003	To develop a regression framework that models multi-state transition or state probabilities directly using pseudo-observations rather than modelling each transition intensity separately	Constructed jackknife pseudo-values from Aalen–Johansen (or Kaplan–Meier) estimators and fitted generalised linear models to these pseudo-values via generalised estimating equations with appropriate link functions	Provides a simple way to regress directly on clinically meaningful state probabilities, avoids specifying full intensity models, works for non-Markov multi-state data, and yields approximately unbiased, easy-to-compute estimators	Relies on working independence assumptions and correct specification of the link function
Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies [6]	2007	To extend the pseudo-value approach for direct regression modelling of state occupation probabilities in multi-state models	Constructed pseudo-values at transition times, estimated covariate effects via GEEs or penalised GEEs with smoothing penalties, and allowed extension to multivariate pseudo-observations for joint modelling of multiple states	Provides direct, interpretable modelling of covariate effects on state probabilities, avoids indirect hazard-based modelling, and allows flexible estimation across multiple states and time points	Requires selection of time points/grid, variance estimation relies on sandwich or bootstrap, and computationally intensive iterative estimation may be needed
Nonparametric regression of state occupation, entry, exit, and waiting times, with multistate right-censored data [95]	2013	To develop nonparametric regression methods for estimating state occupation probabilities and waiting time distributions in progressive multi-state models under censoring, without assuming Markov or semi-Markov structures	Used kernel smoothing of counting and risk processes conditional on covariates, applied inverse probability of censoring weights with Aalen’s linear hazard model	Avoids restrictive parametric/semiparametric assumptions, provides closed-form estimators that are asymptotically valid and robust to censoring misspecification, and offers useful benchmarks for checking parametric models	Performance depends on sample size, and extension to multiple covariates faces dimensionality without additive modelling approaches
Temporal Prediction of Future State Occupation in a Multistate Model from High-Dimensional Baseline Covariates via Pseudo-Value Regression [116]	2017	To develop a method for predicting future state-occupation probabilities in multi-state models using high-dimensional baseline covariates	Combined pseudo-value regression (built from Aalen–Johansen state-occupation estimates) with high-dimensional regression tools such as PLS, SPLS, LASSO, elastic net, and adaptive LASSO	Enables individual-level prediction with thousands of covariates, avoids parametric assumptions, handles censoring via pseudo-values, and performed well in simulations, especially PLS in both sparse and dense settings	Performance deteriorated under high noise-to-signal ratios, depends on correct tuning of high-dimensional models, relies on the quality of pseudo-values, and simulations show sensitivity to heavy censoring
Adjusting for informative cluster size in pseudo-value based regression approaches with clustered time to event data [117]	2023	To extend pseudo-value regression for covariate effects in multi-state models to handle informative cluster size (ICS)	Developed pseudo-values from weighted marginal estimators of SOPs and adjusted for ICS directly in the GEE estimating equations	Provides a robust alternative to Cox-type hazard models by enabling direct regression on SOPs, flexible use of GEE with sandwich variance estimators, and broad applicability to clustered multi-state data	Small cluster numbers can yield unreliable inference; extensions are needed for complex censoring, and bootstrap or resampling strategies may be required
Regression analysis of a future state entry time distribution conditional on a past state occupation in a progressive multistate model [118]	2023	To develop nonparametric estimators and regression methods for future state-entry probabilities and entry-time distributions conditional on past state occupation in progressive multi-state models without Markov or semi-Markov assumptions	Constructed fractional at-risk sets, inverse-probability-of-censoring-weighted competing-risks estimators and combined them with pseudo-value regression to model covariate effects under dependent censoring	Provides valid population-level estimates and covariate effects for future state-entry behaviour in complex progressive systems without relying on Markov assumptions and while allowing dependent censoring	Requires a tree-structured progressive model, accurate modelling of the censoring mechanism, and can become computationally intensive due to recursive fractional-weight calculations
Pseudo-observations for competing risks with	2014	To investigate bias in pseudo-observation-based regression for competing risks when censoring	Replaced standard pseudo-values based on the Aalen–Johansen estimator with pseudo-values derived	Reduces bias in regression coefficients and cumulative incidence estimates under covariate-	Requires correct specification of the censoring model; alternative estimators increase computational

covariate dependent censoring [119]		depends on covariates, and propose modified pseudo-observations to address this issue	from inverse probability of censoring weighted cumulative incidence estimators, where the censoring distribution is modelled (e.g., via a Cox model); evaluated variants that refit or reuse censoring models	dependent censoring; retains the flexibility of pseudo-observation regression	burden or variance; efficiency may be lower compared with standard methods
--	--	---	---	---	--

Most pseudo-value methods construct pseudo-observations from nonparametric estimators of multi-state quantities and then regress these on predictors using methods such as GEE-based approaches, smoothing, weighting procedures, and regularised regression. [6, 95, 114-118] These methods help researchers fit prediction models directly for fixed-horizon risks and accommodate settings such as clustered data or larger baseline covariate sets. This makes them useful when the prediction target is a state occupation probability rather than a transition-specific hazard.

There are several methodological weaknesses to some of these methods. First, pseudo-value regression depends heavily on the quality and stability of the underlying nonparametric estimators, so problems such as censoring, sparse data, or unstable weighting can propagate directly into predictions [6, 95, 115-118]. Second, as pseudo-value models target specific time horizons, performance depends strongly on the choice and spacing of those time points, which can affect both bias and efficiency [6]. Third, although extensions exist for clustered and high-dimensional settings, these can make estimation and uncertainty quantification more difficult, especially when cluster sizes are informative or when censoring depends on many covariates. [116, 117, 119]

Pseudo-value methods are used for multi-state prediction when the primary goal is to model fixed-horizon risks in a clinically interpretable way, but their performance depends on the stability of estimators, design choices regarding censoring adjustment and the selected time points.

Software for pseudo-value-based regression in multi-state settings is available in R. The `pseudo` package computes jackknife pseudo-observations from Aalen-Johansen or Kaplan-Meier estimates for state occupation probabilities and related quantities; these can then be passed into standard regression frameworks. [121] The `mets` package provides an integrated implementation of pseudo-observation regression for multi-state outcomes. [122]

3.4 Model Validation

Validation methods for multi-state prediction models remain limited, with relatively few papers ($n=4$, summarised in Table 5) addressing predictive performance and model fit across different multi-state structures. [123-126] The available literature focuses mainly on discrimination, calibration, and overall prediction error, typically for state occupation or transition probabilities at fixed prediction horizons.

Table 5 Summary of literature on validation methods

Title	Year	Aim	Methods	Benefits	Limitations
Time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks: An application to peritoneal dialysis programme [122]	2016	To evaluate predictive accuracy of nonparametric multi-state regression models using time-dependent ROC methodology in the presence of competing risks	Applied Cox-type structured hazard multi-state regression (STAR) models with flexible nonparametric and time-varying covariate effects; assessed predictive discrimination using time-dependent ROC curves and AUCs across different transitions	Identifies clinically relevant, time-varying covariate effects ,ROC/AUC analysis provides a robust, scale-free metric for comparing discrimination across transitions	Extensions to more flexible hazard structures (e.g., GLASS, P-splines) and alternative ROC definitions remain for future research
Prediction errors for state occupation and transition probabilities in multi-state models [124]	2018	To propose predictive accuracy measures for general multi-state models, covering both baseline and dynamic prediction rules	Developed two estimators for prediction error: one based on inverse probability of censoring weighting (IPCW) and one using pseudo-observations (Aalen–Johansen estimates), with proofs of consistency under specified censoring assumptions	Provides a framework to evaluate and compare predictive performance of multi-state models (including dynamic settings), with practical guidance on when to use IPCW vs. pseudo-observations	Relies on strong assumptions about censoring; bias may occur if censoring distribution is misspecified; standard errors not derived (bootstrap suggested); performance may differ when censoring heavily depends on covariates
Calibration plots for multistate risk predictions models: an overview and simulation comparing novel approaches [123]	2023	To provide methods for evaluating the calibration of transition probabilities in multi-state prediction models	Compared BLR-IPCW, MLR-IPCW, pseudo-value approaches, and Aalen–Johansen estimator using simulations, validation datasets, and clinical examples; recommended combining calibration curves and scatter plots	First detailed framework for calibration of multi-state transition probabilities; demonstrates unbiased estimation under random censoring, identifies method-specific strengths (e.g., pseudo-values most robust to sparse data), and provides practical guidance for applied validation	All methods show bias in low-density regions, performance depends on censoring assumptions and correct use of IPCWs, calibration assessments unstable in small validation cohorts, restricted to transitions out of the initial state and irreversible models
calibmsm: An R package for calibration plots of the transition probabilities in a multistate model [125]	2025	To create software to aid in assessing the calibration of a multistate model	Three approaches: The first two utilised calibration techniques for binary and multinomial logistic regression models in combination with IPCW, the third utilised pseudo-values. All methods are implemented in conjunction with landmarking. Focused on calibration curves, but also allowed estimation of calibration slopes and intercepts	Provides software to assess the calibration for any state in the model, including intermediate states	Further comparison of the approaches in targeted simulations to establish their performance under different censoring mechanisms and assumptions would be valuable

3.4.1 Discrimination

Only one paper addressed discrimination in multi-state prediction, investigating time-dependent ROC curves and AUC measures for competing-risk and multi-state settings. These metrics help model developers assess whether a model appropriately ranks individuals according to their risk of future state occupancy or transition.

3.4.2 Calibration

Most of the identified validation literature focused on calibration and overall accuracy. This included approaches for assessing calibration of multi-state transition probabilities using binary or multinomial IPCW-based regression, pseudo-value methods, and Aalen–Johansen-based benchmarks [124], as well as a broader framework for prediction error for state occupation and transition probabilities using IPCW or pseudo-observations [125]. Extensions have been developed to allow calibration assessment for predictions made from intermediate states [126]. For researchers, these methods are useful because they evaluate whether predicted multi-state probabilities are numerically reliable at clinically relevant time horizons, rather than only whether individuals are ranked correctly.

Most existing validation methods have been developed for simple settings and are only applicable to irreversible models and a small number of fixed horizons, whereas in practice, multi-state models may require validation across multiple states, repeated transitions, and dynamic landmark times. Additionally, uncertainty quantification for multi-state validation metrics remains underdeveloped. Although bootstrap-based intervals are often recommended, work on Aalen–Johansen estimators has shown that commonly used bootstrap procedures may not reproduce the correct limiting behaviour, potentially affecting standard errors and inference unless additional corrections are applied. [125]

Despite validation being a key part of the model pipeline, methods for validating multi-state prediction models are still limited. The present methods provide tools for assessing discrimination, calibration, and prediction error at fixed horizons, but important gaps remain in validating complex multi-state processes and quantifying uncertainty for performance measures across multiple states.

Software for calibrating and validating multi-state prediction models is limited. The `calibmsm` package implements calibration plots for multi-state transition probabilities using BLR-IPCW, MLR-IPCW, pseudo-value and Aalen-Johansen benchmark approaches, and supports calibration from intermediate states. [126]

4 Discussion

This methodological scoping review identified 105 methodological studies on individual-level prediction using multi-state models. We found that most contributions (n=101) focus on developing MSM CPMs, with the majority of those papers focusing on models for Markov approaches, ranging from Cox-type hazards to Bayesian and machine learning frameworks, while semi-Markov, non-Markov and pseudo-value methods remain comparatively underexplored. Most work focused on model specification and fitting, with fewer studies addressing other parts of the prediction pipeline, particularly sample size and validation. Overall, methods are more advanced for modelling transition processes than for developing and assessing multi-state prediction models for routine clinical use.

To our knowledge, this is the first methodological scoping review of methods for producing a clinical prediction model for individual-level predictions across a multi-state process. A

recent review summarised core multi-state model types, including intensity-based, marginal and frailty approaches, and highlighted key challenges such as intermittent observation, dependence structures and the emerging use of machine learning. [127] Our findings align with these themes, particularly the need for stronger validation frameworks and software/open-source code for more complex methods. However, this review is distinct in its systematic mapping of methodological contributions for covariate-specific (individualised) prediction rather than population-based modelling, offering a structured overview of how current approaches address, or fail to address, the unique requirements of risk prediction.

4.1 Methodological Gaps in MSM Prediction Model Development

Despite the added complexity of multiple transitions, states, and time points, no papers addressed sample size for multi-state prediction model development. Existing sample size methods for single-time-to-event prediction models [4, 15] may not transfer directly to multi-state settings, as multi-state models do not target a single endpoint or hazard. Instead, information is split across several transition-specific models or prediction components, often with different event frequencies, follow-up patterns, and censoring mechanisms. A dataset that appears large overall may therefore still provide limited information for particular transitions, especially intermediate or rare transitions, and the correlations between transitions further complicate how effective sample size should be quantified.

Work on missing data specific to multi-state models was also sparse, even though missingness in multi-state settings can affect baseline covariates, time-varying predictors and repeated measurements. Methods for covariate selection and functional form specification were limited, with most approaches assuming linear, time-constant effects despite likely non-linearity and time-varying associations in practice. Throughout model fitting, a recurring challenge was computational scalability, which was particularly limiting for large registry or electronic health record data.

4.2 Methodological Gaps in MSM Prediction Model Validation

Validation methodology was much less developed than model fitting (Table 5). Existing work mainly assessed discrimination, calibration, and prediction error at fixed horizons and often in relatively simple settings. This is poorly aligned with clinical use, where predictions may need to be updated repeatedly across intermediate states and over long follow-up. There was little work on discrimination across multiple landmark times, limited calibration work in data-sparse regions or along full pathways, and underdeveloped methods for uncertainty quantification, particularly for measures based on Aalen–Johansen-type estimands.

4.3 Implications

Future work should prioritise underdeveloped parts of the model development pipeline, especially sample size, missing data, and flexible but interpretable covariate handling. It should also focus on practical usability, including computational efficiency and guidance for implementation in realistic healthcare datasets. Finally, validation should be a core priority, as multi-state prediction models will be clinically useful only if their performance can be assessed reliably across multiple states, time points, and patient pathways.

4.4 Limitations

This review has several limitations. First, the search was restricted to peer-reviewed journal articles written in English, which may have resulted in relevant methodological contributions being missed, particularly those published in non-English literature or disseminated as preprints or technical reports. Second, although search terms were tested against key a priori

references, the breadth of terminology across the multi-state modelling literature means some relevant papers may not have been captured. Third, while the majority of title and abstract screening and all remaining full-text screening was conducted by a single reviewer, high agreement was demonstrated prior to independent screening. Fourth, the grouping of papers into methodological categories required subjective judgement, and some methods span multiple categories; the boundaries between groups should therefore be interpreted with this in mind.

4.5 Conclusion

The methodological literature for multi-state prediction models has progressed substantially in modelling transition processes, but less so in the wider framework needed to develop, validate, and implement these models responsibly.

5 Declarations

5.1.1 *Ethics approval and consent to participate*

Not applicable. This scoping review does not involve human participants, human data, human tissue, or animal subjects.

5.1.2 *Consent for publication*

Not applicable. This manuscript does not contain data from any individual person.

5.1.3 *Availability of data and materials*

The review protocol is prospectively registered and publicly available on the Open Science Framework. Template data collection forms and database-specific search strings for all four databases are available in the supplementary materials. No datasets were generated or analysed during this study beyond the published literature identified through the systematic search.

5.1.4 *Competing interests*

CC, AP, VP, and GPM declare that they have no competing interests.

5.1.5 *Funding*

CC acknowledges receipt of a studentship award from the Health Data Research UK–The Alan Turing Institute Wellcome PhD Programme in Health Data Science (Grant Ref: 218529/Z/19/Z). AP and GPM are supported by the NIHR Better Methods, Better Research grant [Grant Ref: 174468]. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

5.1.6 *Authors' contributions*

CC conceived and designed the review, developed the search strategy, conducted title and abstract screening, performed all full-text screening, extracted data from included studies, and drafted the manuscript. GPM and AP designed the review, contributed to calibration of the title and abstract screening process, reviewed a subset of full-text decisions, and reviewed all drafts of the manuscript. VP contributed supervisory input and reviewed all drafts of the manuscript. All authors read and approved the final manuscript.

5.1.7 *Clinical Trial Number*

Not applicable.

5.1.8 *Acknowledgements*

Not applicable.

References

1. Richard D. Riley, A.P., Paula Dhiman, Lucinda Archer, Glen P. Martin & Gary S. Collins *Clinical prediction models and the multiverse of madness*. BMC Medicine, 2023. **21**.
 2. Julia Hippisley-Cox, C.C., Peter Brindle, *Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study*. BMJ, 2017.
 3. *National Institute for Health and Care Excellence: Guidelines, in Cardiovascular disease: risk assessment and reduction, including lipid modification*. 2023, National Institute for Health and Care Excellence (NICE)
- Copyright © NICE 2023.: London.
4. Martin, G.P., et al., *A decision-analytical perspective on incorporating multiple outcomes in the production of clinical prediction models: defining a taxonomy of risk estimands*. BMC Med, 2025. **23**(1): p. 142.
 5. Putter, H., M. Fiocco, and R.B. Geskus, *Tutorial in biostatistics: competing risks and multi-state models*. Stat Med, 2007. **26**(11): p. 2389-430.
 6. Andersen, P.K. and J.P. Klein, *Regression Analysis for Multistate Models Based on a Pseudo-value Approach, with Applications to Bone Marrow Transplantation Studies*. Scandinavian Journal of Statistics, 2007. **34**(1): p. 3-16.
 7. Martin, G.P., et al., *Toward a framework for the design, implementation, and reporting of methodology scoping reviews*. J Clin Epidemiol, 2020. **127**: p. 191-197.
 8. Andersen, P.K.R., Henrik;, *Models for Multi-State Survival Data*. 2023: Chapman & Hall.
 9. Cornett, C., et al., *Systematic Review of Multi-State Processes for Clinical Prediction*. Open Science Framework, 2025.
 10. Page, M.J., et al., *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*. BMJ, 2021. **372**: p. n71.
 11. Ouzzani, M., et al., *Rayyan-a web and mobile app for systematic reviews*. Syst Rev, 2016. **5**(1): p. 210.
 12. Goldstein, B.A., et al., *Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review*. J Am Med Inform Assoc, 2017. **24**(1): p. 198-208.
 13. Allignol, A., M. Schumacher, and J. Beyersmann, *A Note on Variance Estimation of the Aalen-Johansen Estimator of the Cumulative Incidence Function in Competing Risks, with a View towards Left-Truncated Data*. Biometrical Journal, 2010. **52**(1): p. 126-137.
 14. Chen, B., G.Y. Yi, and R.J. Cook, *Progressive multi-state models for informatively incomplete longitudinal data*. Journal of Statistical Planning and Inference, 2011. **141**(1): p. 80-93.
 15. Riley, R.D., et al., *Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes*. Stat Med, 2019. **38**(7): p. 1276-1296.
 16. Lou, W., et al., *Estimation of Multi-state Models with Missing Covariate Values Based on Observed Data Likelihood*. Commun Stat Theory Methods, 2019. **48**(23): p. 5733-5747.
 17. Curnow, E., et al., *Multiple imputation strategies for missing event times in a multi-state model analysis*. Stat Med, 2024. **43**(6): p. 1238-1255.
 18. Fiocco, M., H. Putter, and H.C. van Houwelingen, *Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models*. Stat Med, 2008. **27**(21): p. 4340-58.
 19. Sennhenn-Reulen, H. and T. Kneib, *Structured fusion lasso penalized multi-state models*. Statistics in Medicine, 2016. **35**(25): p. 4637-4659.
 20. Dang, X., S. Huang, and X. Qian, *Risk Factor Identification in Heterogeneous Disease Progression with L1-Regularized Multi-state Models*. J Healthc Inform Res, 2021. **5**(1): p. 20-53.
 21. Ma, C. and J. Pan, *Multistate analysis of multitype recurrent event and failure time data with event feedbacks in biomarkers*. Scandinavian Journal of Statistics, 2022. **49**(2): p. 864-885.

22. Reeder, H.T., J. Lu, and S. Haneuse, *Penalized estimation of frailty-based illness-death models for semi-competing risks*. Biometrics, 2023. **79**(3): p. 1657-1669.
23. Beck, G.J., *Stochastic Survival Models with Competing Risks and Covariates*. Biometrics, 1979. **35**: p. 427-438.
24. Andersen, P.K., L.P. Hansen, and N. Keiding, *Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous Markov process*. Scandinavian Journal of Statistics, 1991. **18**: p. 153-167.
25. Islam, M.A., *Multistate survival models for transitions and reverse transitions: an application to contraceptive use data*. J R Stat Soc Ser A Stat Soc, 1994. **157**(3): p. 441-55.
26. Alioum, A. and D. Commenges, *MKVPCI: a computer program for Markov models with piecewise constant intensities and covariates*. Comput Methods Programs Biomed, 2001. **64**(2): p. 109-119.
27. Borgan, Ø., *Estimation of covariate-dependent Markov transition probabilities from nested case-control data*. Stat Methods Med Res, 2002. **11**(2): p. 183-202.
28. Christopher H. Jackson, L.D.S., Simon G. Thompson, Stephen W. Duffy, Elisabeth Couto, *Multistate Markov models for disease progression with classification error*. Journal of the Royal Statistical Society: Series D (The Statistician), 2003. **52**(2): p. 131-255.
29. Bhattacharyya, M. and J.P. Klein, *A random effects model for multistate survival analysis with application to bone marrow transplants*. Math Biosci, 2005. **194**(1): p. 37-48.
30. Shu, Y. and J.P. Klein, *Additive Hazards Markov Regression Models Illustrated with Bone Marrow Transplant Data*. Biometrika, 2005. **92**(2): p. 283-301.
31. J. D. Kalbfleisch, J.F.L., *Likelihood analysis of multi-state models for disease incidence and mortality*. Stat Med, 2006. **7**(1).
32. Dantony, E., et al., *Estimating the parameters of multi-state models with time-dependent covariates through likelihood decomposition*. Comput Biol Med, 2016. **69**: p. 37-43.
33. Bluhmki, T., et al., *The wild bootstrap for multivariate Nelson-Aalen estimators*. Lifetime Data Anal, 2019. **25**(1): p. 97-127.
34. Bluhmki, T., et al., *A wild bootstrap approach for the Aalen-Johansen estimator*. Biometrics, 2018. **74**(3): p. 977-985.
35. Dobler, D., M. Pauly, and T. Scheike, *Confidence bands for multiplicative hazards models: Flexible resampling approaches*. Biometrics, 2019. **75**(3): p. 906-916.
36. Beesley, L.J. and J.M.G. Taylor, *EM algorithms for fitting multistate cure models*. Biostatistics, 2019. **20**(3): p. 416-432.
37. Cerquera, L., et al., *A proposal for estimating intensity functions in a three-state model in the presence of arbitrary censoring*. Revista de la Facultad de Ciencias, 2021. **10**: p. 6-19.
38. Ma, C., C. Wang, and J. Pan, *Multistate modeling and structure selection for multitype recurrent events and terminal event data*. Biom J, 2023. **65**(2): p. e2100334.
39. Chen, S., et al., *Bayesian blockwise inference for joint models of longitudinal and multistate data with application to longitudinal multimorbidity analysis*. Stat Methods Med Res, 2024. **33**(11-12): p. 2027-2042.
40. Cremaschi, A., et al., *Seemingly Unrelated Multi-State Processes: A Bayesian Semiparametric Approach*. Bayesian Analysis, 2021.
41. Qin, J. and J. Guan, *Statistical inference of multi-state transition model for longitudinal data with measurement error and heterogeneity*. Communications in Statistics - Theory and Methods, 2023. **53**: p. 1-24.
42. Alafchi, B., et al., *Joint modelling of longitudinal ordinal and multi-state data*. Stat Methods Med Res, 2024. **33**(11-12): p. 1939-1951.
43. You, L.U., et al., *JOINT MODELING OF MULTISTATE AND NONPARAMETRIC MULTIVARIATE LONGITUDINAL DATA*. Ann Appl Stat, 2024. **18**(3): p. 2444-2461.
44. Wang, C., et al., *Bayesian analysis of the Box-Cox transformation model based on left-truncated and right-censored data*. J Appl Stat, 2021. **48**(8): p. 1429-1441.
45. Han, B., et al., *Bayesian approach for flexible modeling of semicompeting risks data*. Statistics in Medicine, 2014. **33**(29): p. 5111-5125.

46. Yen, A.M. and H.H. Chen, *Bayesian measurement-error-driven hidden Markov regression model for calibrating the effect of covariates on multistate outcomes: Application to androgenetic alopecia*. Stat Med, 2018. **37**(21): p. 3125-3146.
47. Kneib, T. and A. Hennerfeind, *Bayesian semi parametric multi-state models*. Statistical Modelling, 2008. **8**(2): p. 169-198.
48. Crowther, M.J. and P.C. Lambert, *Parametric multistate survival models: Flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences*. Statistics in Medicine, 2017. **36**(29): p. 4719-4742.
49. Farewell, V.T., L. Su, and C. Jackson, *Partially hidden multi-state modelling of a prolonged disease state defined by a composite outcome*. Lifetime Data Anal, 2019. **25**(4): p. 696-711.
50. Liu, X., et al., *Modeling Multidimensional Transitions in Health Care*. Sociological Methods & Research, 1997(3): p. 284-317.
51. Huszti, E., et al., *Relative survival multistate Markov model*. Stat Med, 2012. **31**(3): p. 269-86.
52. Ramchandani, R., D.M. Finkelstein, and D.A. Schoenfeld, *Estimation for an accelerated failure time model with intermediate states as auxiliary information*. Lifetime Data Anal, 2020. **26**(1): p. 1-20.
53. Chowdhury, R.I. and M.A. Islam, *Prediction of risks of sequence of events using multistage proportional hazards model: a marginal-conditional modelling approach*. Statistical Methods & Applications, 2020. **29**(1): p. 141-171.
54. Jiang, H., et al., *Classification of disease recurrence using transition likelihoods with expectation-maximization algorithm*. Stat Med, 2022. **41**(23): p. 4697-4715.
55. Lee, K.H., et al., *Bayesian Semi-parametric Analysis of Semi-competing Risks Data: Investigating Hospital Readmission after a Pancreatic Cancer Diagnosis*. J R Stat Soc Ser C Appl Stat, 2015. **64**(2): p. 253-273.
56. Lee, K.H., et al., *Hierarchical models for semi-competing risks data with application to quality of end-of-life care for pancreatic cancer*. J Am Stat Assoc, 2016. **111**(515): p. 1075-1095.
57. Liqueur, B., J.F. Timsit, and V. Rondeau, *Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units*. BMC Med Res Methodol, 2012. **12**: p. 79.
58. O'Keeffe, A.G., L. Su, and V.T. Farewell, *Correlated multistate models for multiple processes: an application to renal disease progression in systemic lupus erythematosus*. J R Stat Soc Ser C Appl Stat, 2018. **67**(4): p. 841-860.
59. Ming-Fang Yen, A. and T. Hsiu-Hsi Chen, *Mixture Multi-state Markov Regression Model*. Journal of Applied Statistics, 2007. **34**(1): p. 11-21.
60. Vasquez, J., et al., *Multistate models with nested frailty for lifetime analysis: Application to bone marrow transplantation recovery patients*. Communications in Statistics - Theory and Methods, 2024. **54**: p. 418 - 436.
61. de Castro, M., M.H. Chen, and Y. Zhang, *Bayesian path specific frailty models for multi-state survival data with applications*. Biometrics, 2015. **71**(3): p. 760-71.
62. Eletti, A., G. Marra, and R. Radice, *A spline-based framework for the flexible modelling of continuously observed multistate survival processes*. Statistical Modelling, 2023. **23**(5-6): p. 495-509.
63. Hatami, F., et al., *A scalable approach for continuous time Markov models with covariates*. Biostatistics, 2024. **25**(3): p. 681-701.
64. Wang, Q., K. Hanewald, and X. Wang, *Multistate health transition modeling using neural networks*. Journal of Risk and Insurance, 2022. **89**(2): p. 475-504.
65. Alsaedi, H., F. Al-Juboori, and R. Kadhim, *Estimating survival rates using artificial intelligence combined with the Aalen-Johansen estimator in multi-state models*. Sustainable Engineering and Innovation, 2025. **8**: p. 1-12.
66. Farcomeni, A. and M. Geraci, *Multistate quantile regression models*. Statistics in Medicine, 2020. **39**(1): p. 45-56.
67. Jackson, C.H., *flexsurv: A Platform for Parametric Survival Modeling in R*. J Stat Softw, 2016. **70**.

68. Therneau, T. *A Package for Survival Analysis in R*. 2026; Available from: <https://CRAN.R-project.org/package=survival>.
69. Rondeau, V., Y. Marzroui, and J.R. Gonzalez, *frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation*. Journal of Statistical Software, 2012. **47**(4): p. 1 - 28.
70. Therneau, T.M. *coxme: Mixed Effects Cox Models*. 2024; Available from: <https://cran.r-project.org/web/packages/coxme/index.html>.
71. Gillaizeau, F., et al., *A multistate additive relative survival semi-Markov model*. Stat Methods Med Res, 2017. **26**(4): p. 1700-1711.
72. Reeder, H.T., et al., *An augmented illness-death model for semi-competing risks with clinically immediate terminal events*. Stat Med, 2024. **43**(21): p. 4194-4211.
73. Wu, S.-C., *A semi-Markov model for survival data with covariates*. Mathematical Biosciences, 1982. **60**(2): p. 197-206.
74. Pons, O., *Estimation in a model for a semi-Markov process with covariates under right-censoring*. Statistics, 2004. **38**: p. 273-293.
75. Dabrowska, D., *Estimation of transition probabilities and bootstrap in a semiparametric markov renewal model*. Journal of Nonparametric Statistics, 1995. **5**(3): p. 237-259.
76. Spitoni, C., M. Verduijn, and H. Putter, *Estimation and asymptotic theory for transition probabilities in Markov renewal multi-state models*. Int J Biostat, 2012. **8**(1): p. 23.
77. Foucher, Y., et al., *A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease*. Biom J, 2005. **47**(6): p. 825-33.
78. Butler, R.W. and D.A. Bronson, *Bootstrap confidence bands for sojourn distributions in multistate semi-Markov models with right censoring*. Biometrika, 2012. **99**(4): p. 959-972.
79. Dabrowska, D.M., *Estimation in a semi-Markov transformation model*. Int J Biostat, 2012. **8**(1): p. Article 15.
80. Eulenburg, C., et al., *A systematic model specification procedure for an illness-death model without recovery*. PLoS One, 2015. **10**(4): p. e0123489.
81. Alaa, A.M. and M.V.D. Schaar, *A hidden absorbing semi-Markov model for informatively censored temporal data: learning and inference*. J. Mach. Learn. Res., 2018. **19**(1): p. 108–169.
82. Richard L. Warr, T.B.W., *Bayesian nonparametric estimation of first passage distributions in semi-Markov processes*. Appl Stochastic Models Bus Ind. , 2020. **36**: p. 237-250.
83. Cadarso-Suárez, C., et al., *Flexible hazard ratio curves for continuous predictors in multi-state models: an application to breast cancer data*. Statistical Modelling, 2010. **10**(3): p. 291-314.
84. Beesley, L.J. and J.M.G. Taylor, *Bayesian variable selection and shrinkage strategies in a complicated modelling setting with missing data: A case study using multistate models*. Statistical Modelling, 2020. **21**(1-2): p. 11-29.
85. Keown-Stoneman, C.D.G., J. Horrocks, and G.A. Darlington, *Weibull multi-state models with misclassification*. Communications in Statistics - Simulation and Computation, 2019. **48**(1): p. 39-57.
86. Król, A. and P. Saint-Pierre, *SemiMarkov: An R Package for Parametric Estimation in Multi-State Semi-Markov Models*. Journal of Statistical Software, 2015. **66**(6): p. 1 - 16.
87. de Wreede, L.C., M. Fiocco, and H. Putter, *mstate: An R Package for the Analysis of Competing Risks and Multi-State Models*. Journal of Statistical Software, 2011. **38**(7): p. 1 - 30.
88. Soutinho, G. and L. Meira-Machado, *Pre-Smoothing Methods for Transition Probabilities in Complex Non-Markovian Multistate Models*. International Statistical Review, 2025. **n/a**(n/a).
89. Cortese, G. and P.K. Andersen, *Competing risks and time-dependent covariates*. Biom J, 2010. **52**(1): p. 138-58.
90. Meira-Machado, L., *Smoothed landmark estimators of the transition probabilities*. SORT-Statistics and Operations Research Transactions, 2016. **40**(2): p. 375-398.
91. Maltzahn, N., et al., *A hybrid landmark Aalen-Johansen estimator for transition probabilities in partially non-Markov multi-state models*. Lifetime Data Anal, 2021. **27**(4): p. 737-760.

92. Soutinho, G. and L. Meira-Machado, *Estimation of the Transition Probabilities in Multi-state Survival Data: New Developments and Practical Recommendations*. WSEAS TRANSACTIONS ON MATHEMATICS, 2020.
93. Titman, A.C., *Transition probability estimates for non-Markov multi-state models*. Biometrics, 2015. **71**(4): p. 1034-41.
94. Meira-Machado, L.U.-Á., J.; Datta, S., *Nonparametric estimation of conditional transition probabilities in a non-Markov illness-death model*. Computational Statistics, 2015. **30**(2): p. 377-397.
95. Mostajabi, F. and S. Datta, *Nonparametric regression of state occupation, entry, exit, and waiting times with multistate right-censored data*. Stat Med, 2013. **32**(17): p. 3006-19.
96. Thomas, H.S. and Z. Mei-Jie, *Direct Modelling of Regression Effects for Transition Probabilities in Multistate Models*. Scandinavian Journal of Statistics, 2007(1): p. 17-32.
97. Andersen, P.K., E.N.S. Wandall, and M. Pohar Perme, *Inference for transition probabilities in non-Markov multi-state models*. Lifetime Data Anal, 2022. **28**(4): p. 585-604.
98. Sun, Y., et al., *DYNAMIC RISK PREDICTION TRIGGERED BY INTERMEDIATE EVENTS USING SURVIVAL TREE ENSEMBLES*. Ann Appl Stat, 2023. **17**(2): p. 1375-1397.
99. Siriwardhana, C., K.B. Kulasekera, and S. Datta, *Flexible semi-parametric regression of state occupational probabilities in a multistate model with right-censored data*. Lifetime Data Anal, 2018. **24**(3): p. 464-491.
100. Azarang, L., T. Scheike, and J. de Uña-Álvarez, *Direct modeling of regression effects for transition probabilities in the progressive illness-death model*. Stat Med, 2017. **36**(12): p. 1964-1976.
101. Lin, H., et al., *A semiparametric transition model with latent traits for longitudinal multistate data*. Biometrics, 2008. **64**(4): p. 1032-42.
102. Datta, S. and G.A. Satten, *Estimation of Integrated Transition Hazards and Stage Occupation Probabilities for Non-Markov Systems Under Dependent Censoring*. Biometrics, 2002. **58**(4): p. 792-802.
103. Huynh, D. and L. Parast. *landpred: Landmark Prediction of a Survival Outcome*. 2025; Available from: <https://cran.r-project.org/web/packages/landpred/index.html>.
104. Kay, R., *The analysis of transition times in multistate stochastic processes using proportional hazard regression models*. Communications in Statistics - Theory and Methods, 1982. **11**(15): p. 1743-1756.
105. Blossfeld, H.-P. and A. Hamerle, *Using Cox Models to Study Multiepisode Processes*. Sociological Methods & Research, 1989. **17**(4): p. 432-448.
106. Huang, Y., *Censored Regression with the Multistate Accelerated Sojourn Times Model*. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 2002. **64**(1): p. 17-29.
107. Van Houwelingen, H.C., *Dynamic Prediction by Landmarking in Event History Analysis*. Scandinavian Journal of Statistics, 2007. **34**(1): p. 70-85.
108. Iacobelli, S. and B. Carstensen, *Multiple time scales in multi-state models*. Statistics in Medicine, 2013. **32**(30): p. 5315-5327.
109. Reulen, H. and T. Kneib, *Boosting multi-state models*. Lifetime Data Anal, 2016. **22**(2): p. 241-62.
110. Putter, H. and C. Spitoni, *Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator*. Stat Methods Med Res, 2018. **27**(7): p. 2081-2092.
111. Soutinho, G. and L. Meira-Machado, *Parametric Landmark Estimation of the Transition Probabilities in Survival Data with Multiple Events*. WSEAS TRANSACTIONS ON MATHEMATICS, 2022.
112. Unseld, T., et al., *Landmarking for Left-Truncated Competing Risk Data*. Biom J, 2024. **66**(8): p. e202400083.
113. Tassistro, E., et al., *Modeling the hazard of transition into the absorbing state in the illness-death model*. Biom J, 2020. **62**(3): p. 836-851.
114. Grand, M.K. and H. Putter, *Regression models for expected length of stay*. Stat Med, 2016. **35**(7): p. 1178-92.

115. Andersen, P., J. Klein, and S. Rosthøj, *Generalised linear models for correlated pseudo-observations, with applications to multi-state models*. Biometrika, 2003. **1**.
116. Dutta, S., S. Datta, and S. Datta, *Temporal Prediction of Future State Occupation in a Multistate Model from High-Dimensional Baseline Covariates via Pseudo-Value Regression*. J Stat Comput Simul, 2017. **87**(7): p. 1363-1378.
117. Anyaso-Samuel, S. and S. Datta, *Adjusting for informative cluster size in pseudo-value-based regression approaches with clustered time to event data*. Stat Med, 2023. **42**(13): p. 2162-2178.
118. Yang, Y., S. Wu, and S. Datta, *Regression analysis of a future state entry time distribution conditional on a past state occupation in a progressive multistate model*. Stat Methods Med Res, 2023. **32**(12): p. 2285-2298.
119. Binder, N., T.A. Gerds, and P.K. Andersen, *Pseudo-observations for competing risks with covariate dependent censoring*. Lifetime Data Anal, 2014. **20**(2): p. 303-15.
120. Per Kragh Andersen , J.P.K., Susanne Rosthøj, *Generalised linear models for correlated pseudo-observations, with applications to multi-state models*. Biometrika, 2003. **90**(1): p. 15-27.
121. Perme, M., M. Gerster, and K. Rodrigues. *pseudo: Computes Pseudo-Observations for Modeling*. 2017; Available from: <https://cran.r-project.org/web/packages/pseudo/index.html>.
122. Holst, K.K. and T. Scheike. *mets: Analysis of Multivariate Event Times*. 2026; Available from: <https://cran.r-project.org/web/packages/mets/index.html>.
123. Teixeira, L., et al., *Time-dependent ROC methodology to evaluate the predictive accuracy of semiparametric multi-state models in the presence of competing risks: An application to peritoneal dialysis programme*. Statistical Modelling, 2016. **16**: p. 409 - 428.
124. Pate, A., et al., *Calibration plots for multistate risk prediction models*. Statistics in Medicine, 2024. **43**(14): p. 2830-2852.
125. Spitoni, C., V. Lammens, and H. Putter, *Prediction errors for state occupation and transition probabilities in multi-state models*. Biometrical Journal, 2018. **60**(1): p. 34-48.
126. Pate, A., et al., *calibmsm: An R package for calibration plots of the transition probabilities in a multistate model*. PLOS ONE, 2025. **20**(6): p. e0320504.
127. Malka Gorfine, R.J.C., Per Kragh Andersen, Terry M. Therneau, Pierre Joly, Hein Putter, Maja Pohar Perme, Michal Abrahamowicz *An Overview and Recent Developments in the Analysis of Multistate Processes*. ArXiv, 2025.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SystematicReviewSupplementaryAug2026.docx](#)