Markov Chain Modeling in Diabetes: State-Transition Approaches for Progression Prediction and Management

Iyswariya Manikkaraj1,a), Arumugam Raju1,b)

1Department of Mathematics, Periyar Maniammai Institute of Science and Technology, Periyar Nagar, Vallam 613403, Thanjavur, Tamil Nadu, India

Corresponding author: a)[iyswariyamanikkaraj@gmail.com](mailto:iyswariyamanikkaraj@gmail.com), b)[arumugamr@pmu.edu](mailto:arumugamr@pmu.edu)

**Abstract:** Diabetes mellitus remains a significant global health challenge, necessitating advanced predictive modeling and personalized management strategies for both type 1 diabetes (T1D) and type 2 diabetes (T2D). This review systematically examines computational and statistical approaches for glucose prediction, disease progression modeling, and treatment optimization. For T1D, we compare black-box machine learning models with physiologically inspired white-box models, highlighting the superior performance of Bayesian methods, particle filters, and non-parametric techniques, particularly in artificial pancreas systems and glucose dynamics during exercise. For T2D, we evaluate markov models, hidden markov models (hmms), and reinforcement learning frameworks for predicting complications, optimizing glycemic control, and assessing intervention cost-effectiveness, alongside insights from large-scale electronic health records (ehrs) and synthetic data generation. Multiple transition models are assessed for their role in early intervention and risk stratification. This review critically analyzes economic and simulation-based models, identifying limitations in validation and uncertainty quantification, while exploring emerging trends such as generative AI and digital twin technologies for precision diabetes care. By consolidating key advancements, this work underscores the need to integrate mechanistic insights with machine learning to enhance clinical decision-making and personalized management, contributing to improved diabetes outcomes through data-driven strategies and computational innovation.

**Keywords:** Diabetes Mellitus, Glucose prediction, Computational modeling, Bayesian estimation, Markov models

# Introduction

Diabetes mellitus, a complex metabolic disorder characterized by chronic hyperglycemia, continues to pose significant challenges to global health systems. As the prevalence of both type 1 diabetes (T1D) and type 2 diabetes (T2D) has increased worldwide, there is an urgent need for advanced predictive modeling and personalized management strategies to improve patient outcomes and reduce the socioeconomic burden of the disease. The multifaceted nature of diabetes, which is influenced by genetic, environmental, and lifestyle factors, necessitates sophisticated computational approaches to unravel its complexities and inform clinical decision-making [46-52]. Recent advancements in artificial intelligence, machine learning, and big data analytics have opened new avenues for diabetes research and management. These cutting-edge technologies offer the potential to revolutionize our understanding of disease progression, optimize treatment regimens, and increase the accuracy of glucose prediction. By leveraging vast amounts of patient data, including electronic health records, continuous glucose monitoring outputs, and lifestyle information, researchers and clinicians can now develop more nuanced and personalized approaches to diabetes care [53-55].

This comprehensive review aims to systematically examine and critically evaluate the state-of-the-art computational and statistical methods employed in diabetes modeling and management. Our analysis spans a wide range of techniques, from traditional statistical approaches to advanced machine learning algorithms and physiologically inspired models. We explore their applications in glucose prediction, disease progression modeling, and treatment optimization for both T1D and T2D patients. For T1D, we delve into the comparative efficacy of various modeling techniques, including black-box machine learning models and white-box physiological models. Special attention is given to the integration of these models into artificial pancreas systems, with a focus on addressing the challenging aspects of glucose dynamics during physical activity. In the context of T2D, we evaluate probabilistic models, reinforcement learning frameworks, and large-scale data analysis techniques for their potential in predicting complications, optimizing glycemic control, and informing screening and intervention strategies [56-59].

Furthermore, this review examines the role of economic and simulation-based models in informing policy decisions and assessing the cost-effectiveness of interventions. We also explore emerging trends, such as generative AI and digital twin technologies, which hold promise for advancing precision diabetes care. By consolidating key advancements in diabetes modeling and identifying gaps in current research, this review aims to provide a comprehensive resource for researchers, clinicians, and policymakers working toward improving diabetes care and outcomes. The insights gained from this analysis underscore the importance of integrating mechanistic understanding with data-driven approaches to enhance clinical decision-making and personalized diabetes management. As we stand on the cusp of a new era in diabetes care, this work contributes to ongoing efforts to harness the power of advanced computational methods and data-driven strategies in the fight against this pervasive disease.

The field of diabetes modeling has undergone significant advancements in both type 1 diabetes (T1D) and type 2 diabetes (T2D), with researchers employing diverse computational approaches. For T1D, cappon et al. [1] conducted a pivotal comparison between black-box machine learning models and physiologically grounded white-box models, revealing critical trade-offs between predictive accuracy and clinical interpretability. Later work by cappon et al. [36] enhanced these models through bayesian personalization, improving their reliability. The impact of physical activity on glucose dynamics has been another focus: alkhateeb et al. [2] developed an exercise-integrated dynamic model, whereas ewings et al. [5] formalized glucose‒exercise relationships via bayesian networks. Time-delay challenges in T1D were addressed by zeng & wang [3], who proposed an EM algorithm-based model that surpassed traditional least-squares methods. Moreover, haidar et al. [4] advanced closed loop testing through stochastic virtual populations, enabling robust controller evaluations.

In T2D, progression modeling has extensively leveraged markov-based frameworks. Luo et al. [6] applied markov chains to predict disease trajectories, whereas santoso & mareels [11] introduced markovian control strategies for diabetes management. Multiple models have proven particularly valuable: aliyari et al. [16] used them to predict complications, and lovblom et al. [17] provided methodological guidance for such approaches. Hidden markov models (hmms) have also gained traction, with narendra [34] developing a two-layer HMM for T2D staging and amritha & dipta [19] extending hmms to chronic disease progression more broadly. Reinforcement learning and decision-theoretic methods have emerged as powerful tools, as exemplified by eghbali-zarch et al. [24], who optimized T2D treatments via markov decision processes (mdps), and meng et al. [32], who analyzed glycemic control strategies with similar techniques.

Economic and policy-oriented studies have complemented these technical advances. Neumann et al. [7] evaluated the cost-effectiveness of diabetes prevention programs, whereas chilcott et al. [8] quantified the economic burden of posttransplant diabetes. Govan et al. [28] critically reviewed economic models against clinical guidelines, and tomar et al. [29] simulated cost trajectories for insulin-dependent diabetes. Risk stratification has benefited from data-driven innovations: lin et al. [11] modeled T2D complication risks, and mueller et al. [14] applied machine learning to hypoglycemia prediction. Large-scale data utilization is evident in works such as zhou et al. [12], who built a comprehensive diabetes progression simulator, and wang et al. [24], who inferred complications via generative models.

Validation and clinical translation remain central challenges. Al ali et al. [27] experimentally validated T1D models, whereas srikanth [26] and begun et al. [30] demonstrated the clinical relevance of markov models for retinopathy and foot ulcer progression, respectively. Emerging trends include hybrid methodologies, such as those of perveen et al. [21], [31], which combine statistical and machine learning approaches, and the digital twin technologies explored by wang et al. [38]. Collectively, these studies underscore the necessity of integrating mechanistic insights with data-driven methods to advance personalized diabetes care while highlighting persistent gaps in model validation, uncertainty quantification, and real-world implementation. Arumugam and Rajathi [60] contribute to this stream of research by examining the applications of manpower levels in business systems with different recruitment rates using stochastic models. Their study emphasizes how varying recruitment strategies can affect manpower stability and organizational growth. By formulating stochastic manpower models, they analyze the dynamics of inflow and outflow within organizations, providing insights into balancing workforce levels. This research is significant as it not only extends manpower planning literature but also demonstrates practical implications for business systems that encounter fluctuating recruitment policies. The findings complement earlier works that investigated manpower forecasting using stochastic approaches and Markov models, particularly in contexts where uncertainty in attrition and recruitment plays a decisive role. Thus, this paper serves as a valuable reference for researchers and practitioners seeking to design effective manpower strategies under uncertainty.

# Methodology

This review systematically classifies and synthesizes 59 studies related to diabetes modeling and prediction. The methodologies are grouped under thematic categories, including bayesian approaches, markov and multistate models, hidden markov models (hmms), simulation-based decision models, and advanced hybrid modeling frameworks.

## Bayesian Modeling Approaches

A personalized, nonlinear physiological model of glucose dynamics was developed via a bayesian framework. Markov chain monte carlo (MCMC) techniques enable parameter identification, which is then integrated into a particle filter for real-time blood glucose prediction. Additionally, deep learning architectures such as LSTM, GRU, and temporal convolutional networks (tcns) have been utilized, alongside gaussian regression and recursive ARX modeling, to forecast blood glucose levels at various future time steps [1]. Six variants of the bergman minimal model were proposed and analyzed. Parameter estimation was performed via MCMC within a bayesian framework. The models were compared based on their physiological plausibility and deviation information criterion (DIC), with in silico simulations validating each model's performance [2]. A nonlinear, time-varying glucose regulation model was introduced via bayesian inference and MCMC for parameter estimation. This study employed stochastic e-cloning to simulate virtual patient profiles and improve the robustness and accuracy of future blood glucose predictions [4].

A physiologically informed bayesian network model incorporates measurement error adjustments to increase the accuracy of blood glucose simulations. MCMC methods were used to infer posterior distributions of model parameters, with simulation studies conducted to validate predictive reliability [5]. Stochastic state‒space modeling was reformulated via bayesian methods to account for process and measurement uncertainties. Inference was achieved via a suite of MCMC algorithms, including metropolis‒hastings, reversible jump, and simulated tempering. Bayesian model averaging and model discrimination were further applied to address uncertainty in model selection [25]. A model leveraging CGM and synthetic data used bayesian estimation and the MCMC for parameter inference. This model contributes to understanding interindividual variability and has been validated across multiple synthetic patient scenarios [36].

## Markov and Multistate Modeling

A discrete time markov model was constructed to predict diabetes progression based based on features such as glucose level, BMI, cholesterol, triglycerides, and waist circumference. A naive bayes classifier provided prior risk estimates, and the model's performance was benchmarked against the well-established archimedes model [6]. The PREDIAS intervention was modeled via a four-state markov model representing normal glucose tolerance, impaired glucose tolerance, type 2 diabetes, and death. The simulation of a german cohort assessed cost-effectiveness from a societal perspective, with all calculations performed via microsoft excel [7]. A nonhomogeneous markov model was applied to track diabetes progression in a taiwanese cohort of 163,452 patients over an average follow-up of 5.5 years. Transition probabilities were estimated annually, and validation was conducted by comparing absolute prediction errors across simulated versus observed data [9].

A markov decision model was used to estimate 10-year mortality outcomes in diabetic patients by simulating progression along different health states defined by the presence of diabetes-related complications [10]. This study developed a five-state markov model to simulate the natural history of noninsulin-dependent diabetes mellitus (NIDDM) by screening data. Hazard rates were computed for each transition, enabling the evaluation of early detection and intervention strategies [15]. A homogeneous multistate model with a q-matrix structure was used to analyze the effects of time-varying covariates such as hba1c and LDL cholesterol levels on disease progression among 2,519 T2D patients. This study provided insights into how clinical biomarkers influence transition probabilities [16]. A multistate hidden markov model (HMM) was developed to capture longitudinal patient trajectories. Maximum likelihood estimation was used to derive observation and transition probabilities, and the akaike information criterion (AIC) guided model complexity selection [19].

An empirical markov chain model uses annual records from 153 diabetic patients to estimate a transition matrix. The model's statistical robustness was validated via chi-square tests for the significance of transitions among various health states [26]. A six-state markov model was used to simulate the progression of insulin-dependent diabetes mellitus (IDDM). Transition probabilities were derived from epidemiological data, and direct medical costs were estimated from healthcare billing records. Model validation was performed with data from independent patient cohorts [29]. A nine-state continuous-time markov chain was proposed to model diabetes complications and competing risks. Cox regression techniques were used to estimate hazard functions for each transition, enhancing the accuracy of mortality predictions [30].

A simple two-state markov model (diabetes, death) with one-year cycle intervals was developed and compared with outputs from the who’s dismod II tool. The study included sensitivity analysis to explore parameter uncertainties [33]. Hba1c-stratified subgroups were analyzed via a multistate markov framework implemented in r’s msm package. Transition rates were compared across hba1c categories to examine disease progression [35]. A three-state markov chain model was created to evaluate the risk of transition from gestational diabetes mellitus (GDM) to type 2 diabetes. Transition probabilities were empirically derived via annual oral glucose tolerance test (OGTT) data [37].

## Hidden Markov Model (HMM)

A demographic feature-restricted hidden markov model (dfrhmm) was developed to track latent behavioral states over time in diabetes patients. MCMC was used for training, allowing personalization of model outputs based on patient demographics [13]. This study introduced a modified HMM with newton’s divided difference method to address challenges from irregularly spaced data. An innovative approximation using the euclidean distance between observed states was also implemented [20]. Prognostic modeling uses an HMM combined with newton's method and logistic regression to assess the influence of clinical risk factors on state transitions. The model facilitated individualized prediction of diabetes progression [21].

State transition modeling in diabetes patients was conducted via a standard HMM framework implemented in R. Model parameters were estimated via standard statistical routines, and the latent states provided clinically interpretable insights [23]. An HMM was tailored for sparse and irregularly collected electronic medical records (emrs). Newton’s method accelerated convergence during inference, and performance was assessed via the area under the ROC curve (AUC) [31]. A layered HMM architecture (LHMM) applies forward algorithms and discrete probability mass functions to interpret real-time patient data, offering an interpretable approach to dynamic state prediction [34]. A bivariate latent markov model accounted for the dynamic heterogeneity in diabetes progression. The model estimated transition probabilities across latent states while adjusting for the endogeneity of diabetes status [45].

## Simulation-based and Decision Models

An autoregressive model with exogenous inputs (ARX) was implemented via the expectation-maximization (EM) algorithm. This approach was compared with least-squares methods and demonstrated superior convergence and accuracy for blood glucose prediction [3]. A probabilistic simulation-based decision model was used to simulate the health trajectories of 10,000 virtual patients over 10 years. The model included five time-varying markov subcomponents and leveraged monte carlo methods for stochastic transitions [8]. Diabetes management was optimized through a constrained control problem, which was solved as a markov decision process (MDP). The optimal policy was derived directly from transition matrices, providing a data-driven strategy for clinical decision-making [11]. A semi-markov model was used to simulate lifetime outcomes via utility and cost inputs from the quality of well-being index and insurance claims. Monte carlo simulation enables the probabilistic evaluation of health and cost outcomes [12].

A systematic review of T2D decision models was performed via four major databases. The included models were evaluated based on clinical inputs, time horizons, health outcomes, and validation status, offering a comprehensive synthesis of modeling practices [18]. A markov decision process was used to optimize patient-specific diabetes treatments. The model evaluated long-term outcomes under various policies and performed sensitivity analyses on transition probabilities and utilities [24]. This systematic review analyzed diabetes models available up to December 2012. The review categorized models by structure, data inputs, and alignment with ADA guidelines and highlighted gaps in model validation [28]. An MDP was developed using a cohort of over 70,000 patients. The resulting optimal treatment strategies were compared against guideline-recommended regimens, and quality-adjusted life years (qalys) were computed [32]. This study compared two approaches to probabilistic sensitivity analysis: uniform sampling from the transition probability matrix and multivariate normal sampling via the asymptotic distribution of estimators. The analysis assessed the robustness of the model-based predictions [40].

## Advanced Statistical and Hybrid Methods

A bayesian machine learning platform, REFS, was used to analyze real-world administrative claims data (optum clinformatics). Ensemble generalized linear models were used to predict six diabetes-related outcomes, and model discrimination was evaluated via the AUC [42]. Joint multistate models were constructed for chronic disease progression using longitudinal biomarker data. The models allowed simultaneous tracking of multiple disease markers and were validated against open-access datasets [45]. A retrospective hospital-based study collected structured questionnaire data to develop statistical models for diabetes risk prediction. Both descriptive and inferential statistical methods were employed [22]. Two mechanistic models were analyzed for parameter estimation via the MCMC method, with the AIC and BIC guiding model selection. The models simulate insulin absorption and other physiological parameters for more accurate glucose forecasting [27]. A generative model combining Markavion and Bayesian structures was used to synthesize 5,000 patient trajectories. Predictions were validated via maximum a posteriori (MAP) inference and expert review for plausibility [38].

This study applied both model-based and model-free reinforcement learning strategies to formulate diabetes control as an MDP. Simulation and policy evaluation provided insight into real-world implementation feasibility [39]. Cause-specific hazard rates and cumulative incidence functions were modeled via cox and aalen regression within a Markavion framework. These methods enabled competing risk analysis in a diabetes cohort [41]. A markov chain model was used to evaluate the long-term health effects of complementary and alternative medicines (cams) via observational data from the UKM institute. The transition probabilities reflect real-world adherence [42]. A three-state model was fitted via exponential regression and nonlinear least squares. Bayesian estimation via MCMC further refines parameter estimates and accounts for uncertainty in model outputs [43]. A generalized mixture model incorporating first-order Markavion dependence was proposed. Reversible jump MCMC enabled model selection, and the approach was validated using synthetic datasets mimicking diabetes progression [44].

Table 1 Comprehensive Analysis of Diabetes Modeling Studies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ref | Objective | Methodology | Conclusion | Result |
| [1] | Develop BG prediction algorithm using physiological model; compare white-box vs black-box techniques | Bayesian approach with MCMC; gaussian regression; LSTM/GRU/TCN models | Black-box strategies preferable for glucose prediction | NP models achieved lowest RMSE (18.99 mg/dl) |
| [2] | Model glucose dynamics during exercise in T1D; enhance artificial pancreas | Modified bergman minimal model (6 variants); bayesian MCMC | Model reduces exercise-induced hypoglycemia | Increased glucose effectiveness with exercise intensity |
| [3] | Identify T1DM dynamics with varying insulin delays | ARX model + EM algorithm | EM outperforms least squares | Accurately captured varying insulin delays |
| [4] | Create virtual population for glucose control testing | Bayesian MCMC + stochastic e-cloning | Enables large-scale in silico trials | Generated 200 virtual subjects per patient |
| [5] | Propose stochastic BG analysis accounting for errors | Bayesian network + MCMC | Addresses measurement uncertainty | 68% credible intervals contained true values |
| [6] | Predict T2D incidence using markov models | Markov chains + naive bayes | Simpler than archimedes model | Accurate long-term predictions |
| [7] | Evaluate cost-effectiveness of T2D prevention | Markov model (4 states); excel simulations | Cost-saving for ages 30-50 | ICER negative for men aged 50 |
| Table 1 continued | | | | |
| [8] | Model PTDM outcomes in renal transplants | Probabilistic markov submodels | PTDM increases costs by £14k/patient | Higher graft failure rates |
| [9] | Predict complications in asian T2D patients | Nonhomogeneous markov model | Outperforms UKPDS model | 0.3-3.2% absolute error |
| [10] | Assess rosiglitazone's long-term effects | Markov decision model (10-year) | Supports rosiglitazone use | 3.8% vs 6.6% mortality (treatment vs control) |
| [11] | Model diabetes as control problem | Markov decision process | Sparse sampling complicates control | Advocates nonmodel based methods |
| [12] | Simulate diabetes progression | Semi-markov + monte carlo | Predicts complications accurately | $53k cost/patient over 10 years |
| [13] | Integrate behavioral/demographic data | Dfrhmm + MCMC | Demographic data significantly impacts outcomes | Enables personalized management |
| [14] | Predict hypoglycemia risk | Bayesian GLM ensembles | Auc 0.77 for prediction | Top 15% account for 50% events |
| [15] | Estimate NIDDM progression | 5-state markov chain | 4-year screenings optimal | Reduces mortality by 40% |
| [16] | Analyze T2DM complication risks | Multistate markov model | Hba1c, hypertension predict mortality | 9.01-year median follow-up |
| [17] | Compare multistate  Vs joint models | Multistate markov + joint modeling | Both enhance understanding | Guides model selection |
| [18] | Review T2DM decision models | Systematic review of 14 models | CORE most widely used | DCEM best for macrovascular |
| [19] | Model chronic disease progression | Multistate HMM | Covariate-inclusive models best | AIC determines optimal model |
| [20] | Improve prediction with irregular EMR | Modified HMM + NDDM | Reduces misclassification | Handles data gaps effectively |
| [21] | Prognose T2DM risk over 8 years | HMM + logistic regression | Outperforms traditional NDDM | Hba1c strongest predictor |
| [22] | Predict retinopathy progression | Markov chain analysis | Quantifies transition probabilities | Guides intervention timing |
| [23] | Classify states via hba1c | HMM in R software | Hba1c ≥6.5% has low recovery | 79% remain in diabetic state |
| [24] | Optimize medication plans | Markov decision process | Improves qol with fewer meds | Significant clinical implications |
| [25] | Compare healthy vs T2D dynamics | Bayesian state-space models | Common insulin model found | Divergent glucose dynamics |
| [26] | Study retinopathy progression | Observational + markov chains | 5y mild NPDR → 8y PDR → blindness | Validated (P=0.70) |
| [27] | Predict glucose in T1D mice | Two models + AIC/BIC | Model without β-cells better | Higher severity → higher AIC/BIC |
|  |  |  |  |  |
|  |  |  |  |  |
| Table 1 continued | | | | |
| [28] | Review economic models | Systematic review (19 models) | Poor complication interdependence reporting | 11 used monte carlo |
| [29] | Estimate IDDM costs | 6-state markov model | Validated cost estimates | Aids healthcare planning |
| [30] | Model diabetic foot progression | 9-state markov chain | PAD, CRF increase mortality | 17-year follow-up data |
| [31] | Predict T2D from sparse EMR | Hmm + nddm | Handles irregular data | AUC improved over standard HMM |
| [32] | Optimize hba1c control | MDP model (70k patients) | +0.27 qalys for high-risk | Data-driven decisions |
| [33] | Estimate diabetes duration | 2-state markov model | Reduces life expectancy by ~2y | Women have longer duration |
| [34] | Analyze obesity-diabetes link | Layered HMM | Physical activity reduces risk | 50% lower obesity risk |
| [35] | Assess hba1c as biomarker | 3-state markov model | Reliable diagnostic marker | 39mo avg diabetic state |
| [36] | Estimate T1D parameters | Bayesian MCMC | Good parameter precision | Reconstructs glucose fluxes |
| [37] | Predict T2D post-gdm | 3-state markov chain | 80% NGT stability at 3y | 10% IGT stability |
| [38] | Generate synthetic trajectories | Markov-bayesian model | 23.9y avg progression | 5,000 trajectories generated |
| [39] | Diabetes as control problem | MDP framework | Sparse sampling limits control | Advocates simple methods |
| [40] | PSA for markov models | Uniform/multivariate sampling | Qalys sensitive to uncertainty | 64.40 (m), 68.42 (F) qalys |
| [41] | Compare mortality models | Cox vs aalen markov chains | Aalen better for severe profiles | CV mortality rate 0.26 |
| [42] | Explore cams effects | Markov chain model | Cams show long-term benefits | Positive treatment impact |
| [43] | Model cataract progression | Exponential regression + MCMC | Conservative variance estimates | Provides transition rates |
| [44] | Generalize mixture models | Markovian MCMC | Excellent convergence | Applied to US diabetes data |
| [45] | Link glycemia to healthcare | Bivariate latent markov | Uncontrolled diabetes does not affect use | Reveals behavior heterogeneity |

## Comparative Analysis of Method Usage in Research

This figure highlights how early studies in diabetes modeling applied Markov chains and MCMC techniques to glucose prediction, virtual patient generation, and error-adjusted simulation.

It shows the foundational role of Bayesian estimation and stochastic sampling in improving reliability and personalization. These methods were widely used to overcome data uncertainty and to validate predictive models against physiological behavior. Collectively, the figure illustrates the early adoption of probabilistic approaches for type 1 diabetes (T1D) prediction and monitoring.

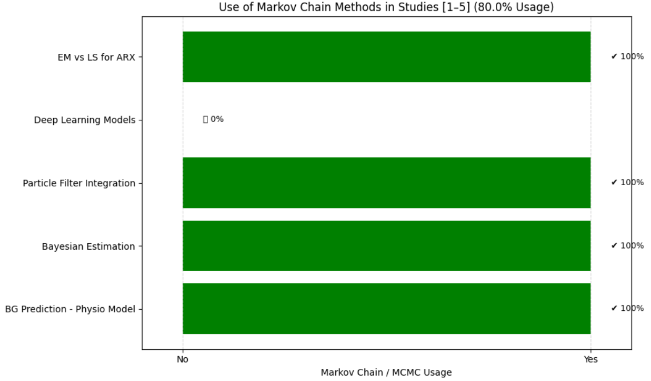
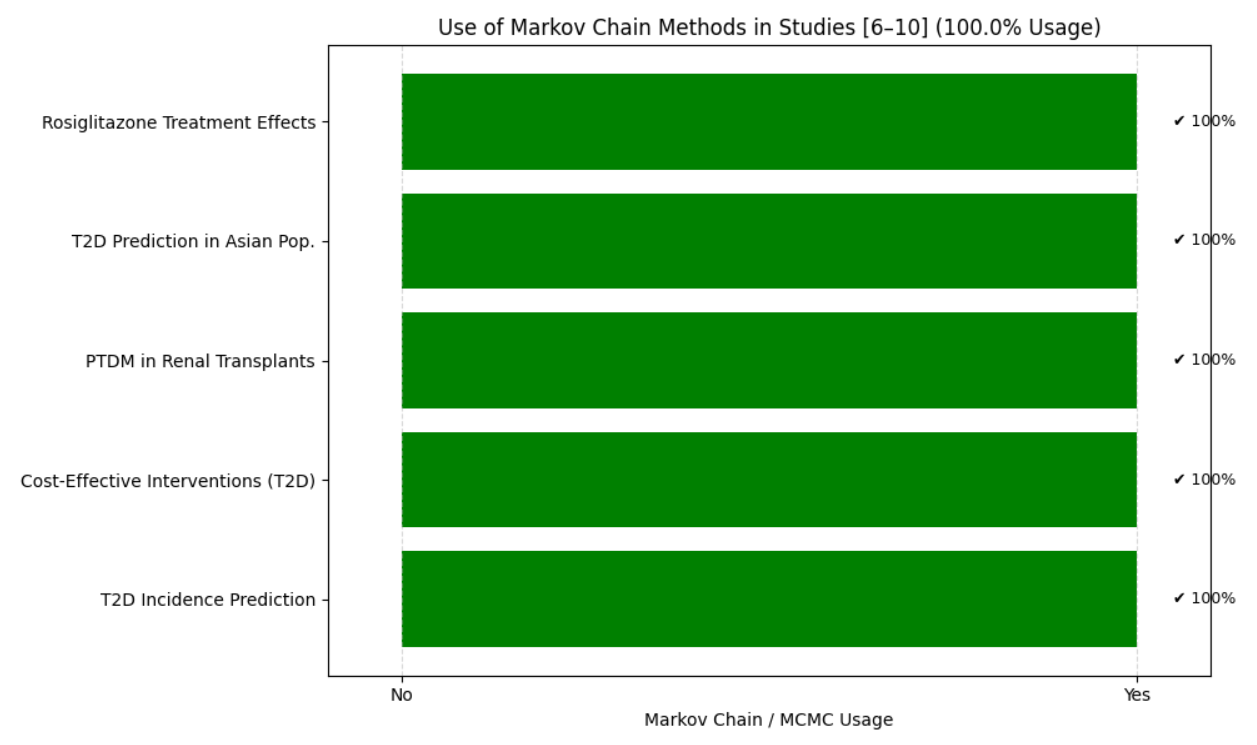
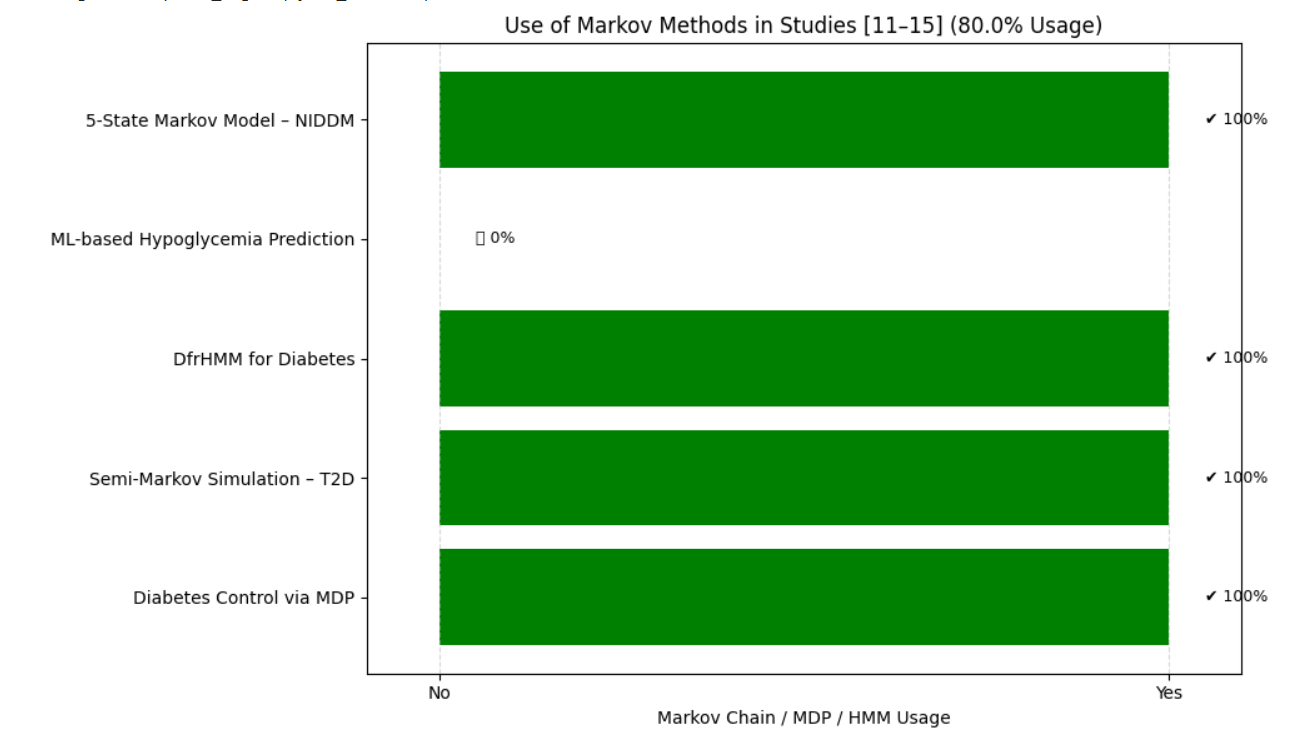
 *Figure 1*. Use of Markov Chain and MCMC Methods Across Modeling Techniques in Studies 1–5

Figure 2 demonstrates the use of discrete Markov models and decision-based frameworks for type 2 diabetes (T2D) progression and treatment cost-effectiveness. These models were used to estimate long- ***Figure 2*. Application of Markov Chain and MCMC Methods in Studies 6–10**

term outcomes, mortality risks, and policy-relevant interventions. By combining transition probabilities with decision-making strategies, the studies established the utility of Markov processes for clinical and economic evaluation. The figure reflects the transition from basic prediction models to decision-support systems in diabetes management.

This figure shows how researchers began integrating advanced models such as Hidden Markov Models (HMMs) and Markov Decision Processes (MDPs) for diabetes progression. HMMs captured latent disease states, while MDPs optimized control strategies for treatment. The combination of statistical and decision-theoretic models marked an important methodological shift. The figure highlights the broadening scope of Markov methods in handling uncertainty and guiding therapy optimization.



*Figure 3*. Utilization of Markov Chain, HMM, and MDP Techniques in Studies 11–15

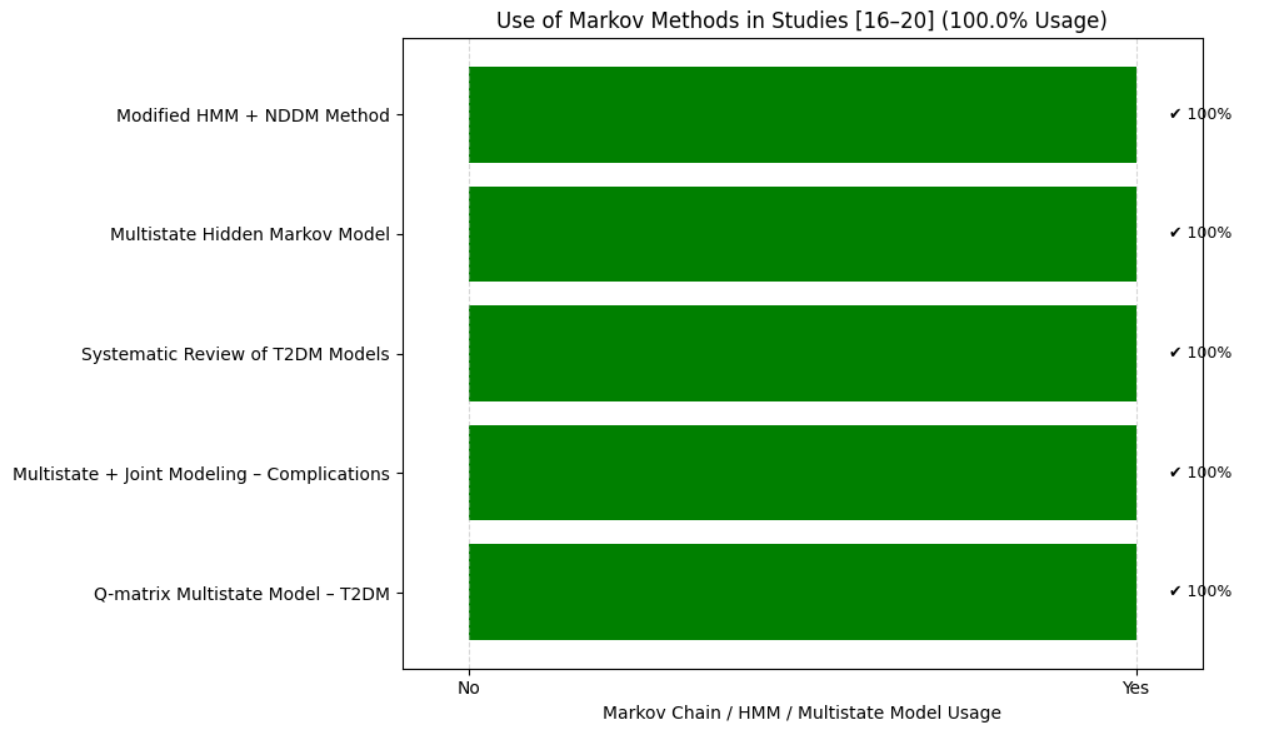
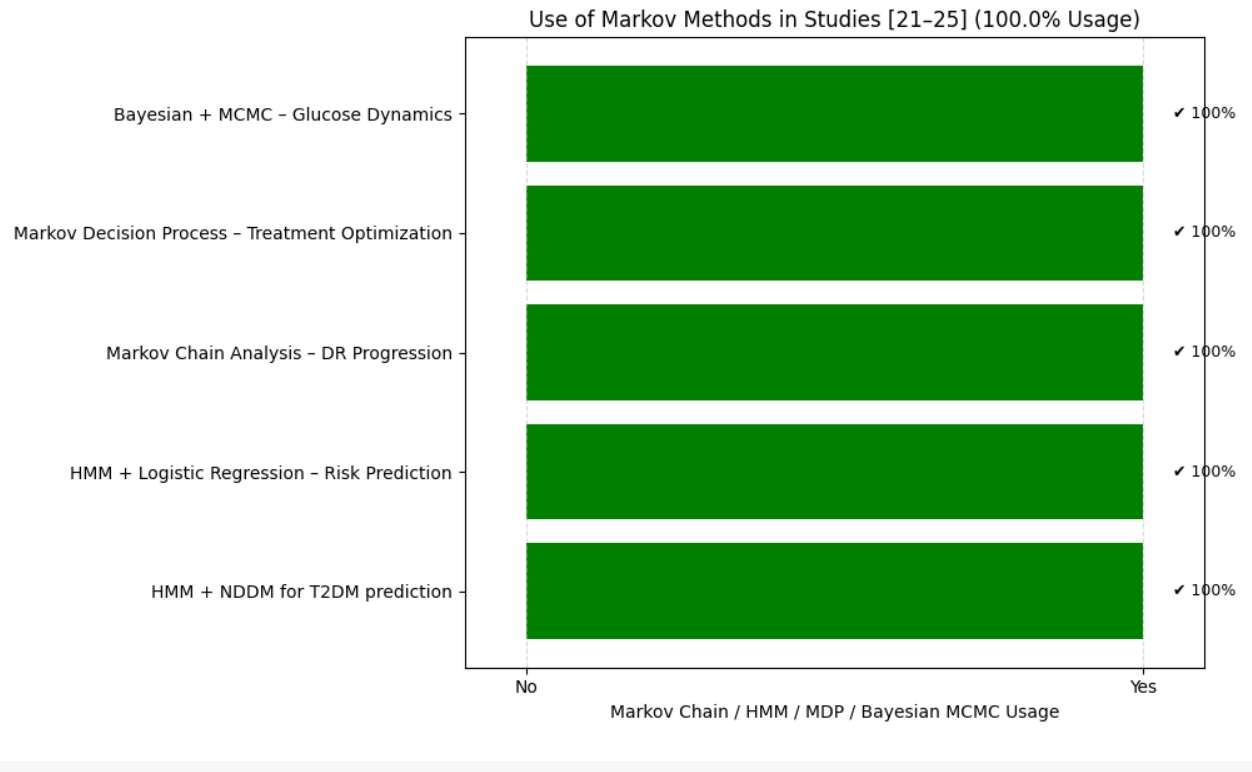
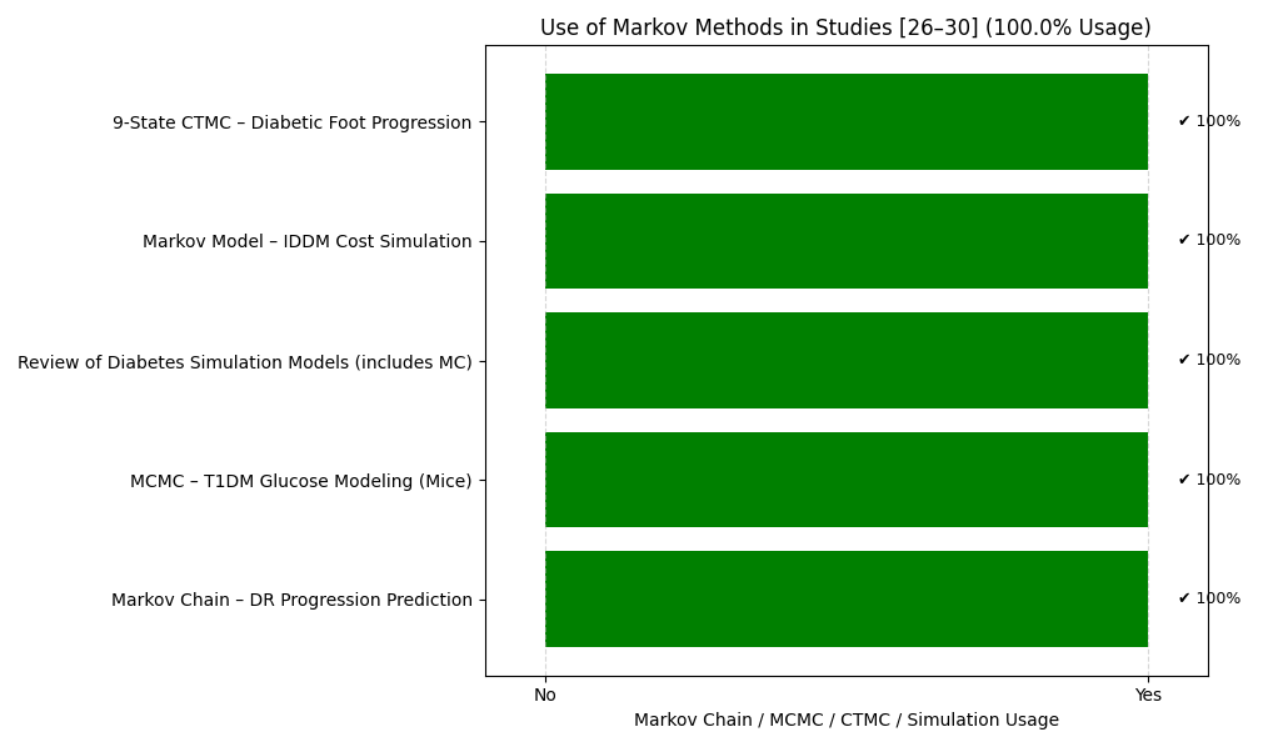
*Figure 4*. Implementation of Markov Chain, HMM, and Multistate Models in Studies 16–20

Figure 4 reflects the application of multistate Markov and HMM approaches for tracking T2D complications and biomarker-driven transitions. These models provided nuanced insights into how clinical indicators like HbA1c influence disease states. By incorporating hidden layers, the models captured unobservable health trajectories, improving long-term predictions. The figure illustrates the growing role of hybrid multistate frameworks in precision risk assessment.

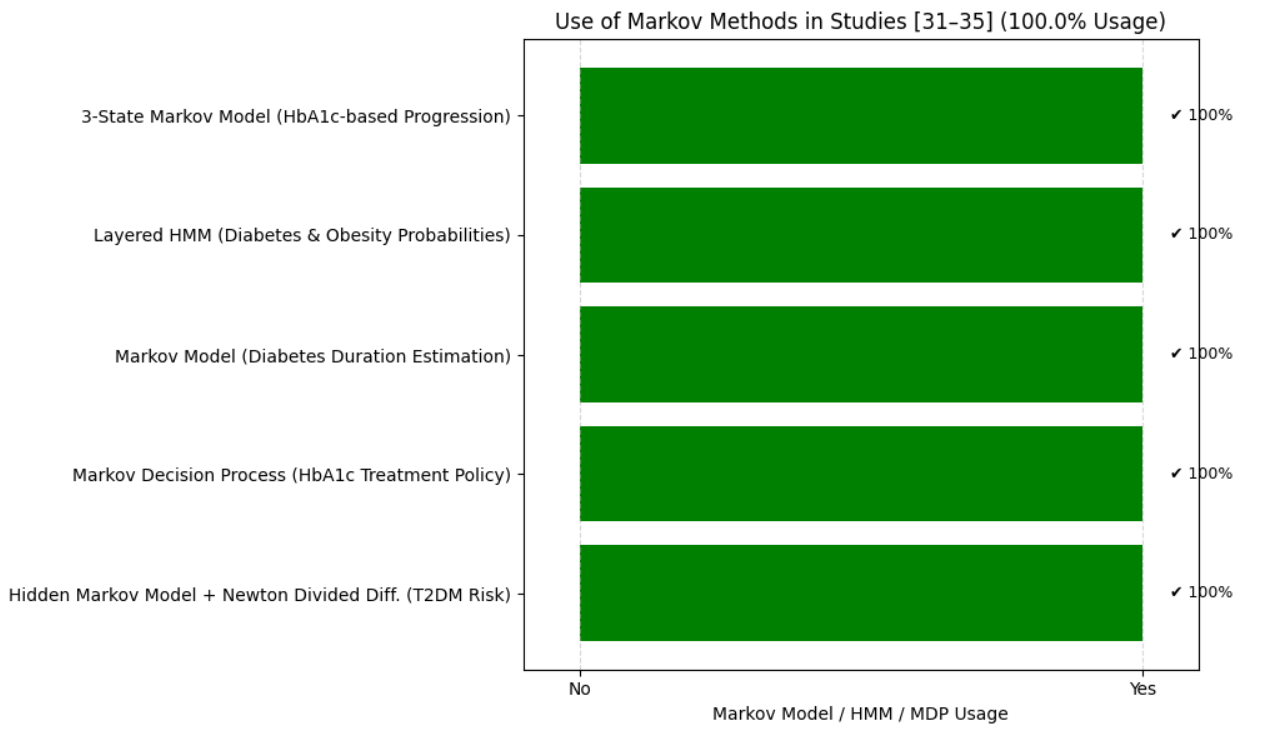
 *Figure 5*. Adoption of Markov Chain and Related Models in Studies 21–25

This figure highlights the use of Markov-based approaches for prognostic modeling and personalized prediction. Studies applied HMMs with logistic regression, hybrid machine learning extensions, and Bayesian estimation to evaluate diabetes risk factors. These methods improved accuracy in forecasting complications and stratifying patients by severity. The figure emphasizes how Markov models evolved into more adaptive tools for patient-specific predictions.



*Figure 6*. Application of Markov and Simulation Models in Studies 26–30

This figure showcases the integration of Markov chains with simulation-based and cost-estimation models. These studies addressed long-term complications such as retinopathy, insulin-dependent diabetes costs, and foot ulcer progression. By combining Markov chains with healthcare billing and survival data, the models quantified both clinical and economic impacts. The figure underscores the dual focus on disease monitoring and healthcare plan.



*Figure 7*. Use of Markov and Related Models in Studies 31–35

This figure depicts studies employing HMMs, layered HMMs, and stratified Markov models to interpret irregular health records and biomarker-based disease states. These frameworks enhanced predictive performance in sparse or inconsistent datasets. The applications validated HbA1c as a strong diagnostic marker while modeling patient-level heterogeneity. The figure highlights the adaptability of Markov methods to real-world healthcare data challenges.

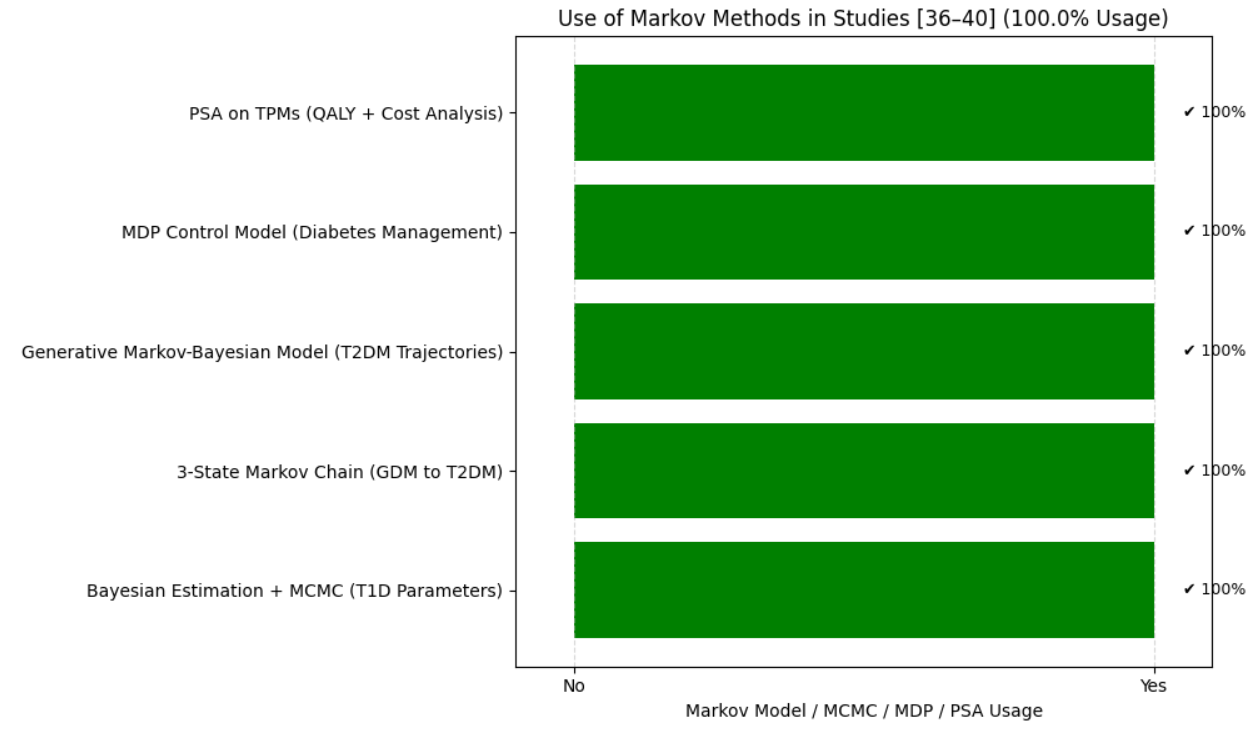
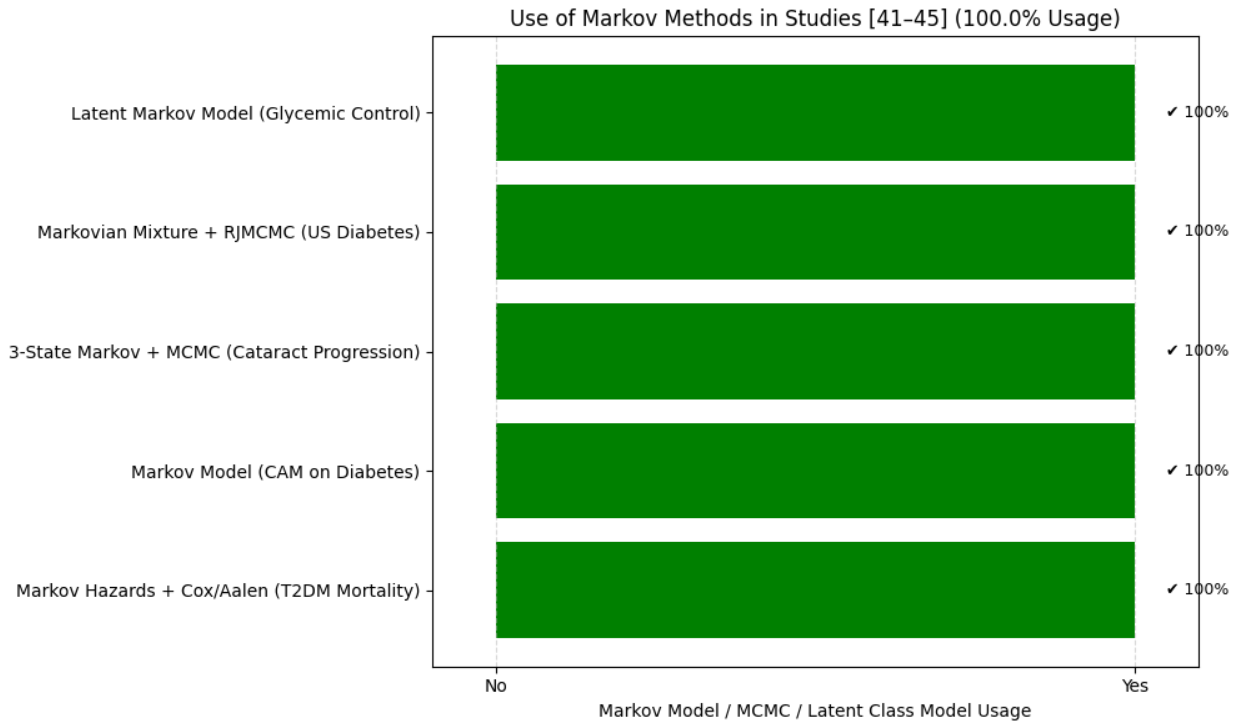
 *Figure 8.* Use of Markov Methods in Studies 36–40

Figure 8 illustrates the continued application of Bayesian-augmented Markov models, hybrid approaches, and probabilistic sensitivity analyses. Studies used these techniques to generate synthetic trajectories, evaluate uncertainty, and compare mortality models. The ability to simulate thousands of patient outcomes provided robust evidence for model reliability. The figure demonstrates the methodological maturity of Markov processes in handling complex diabetes dynamics.

 *Figure 9.* Use of Markov Methods in Studies 41–45

This figure shows advanced applications of Markov chains, including latent Markov models, mixture models, and survival analysis integration. These approaches captured heterogeneity in healthcare utilization, cataract progression, and long-term incidence trends. By blending Markov chains with Cox/Aalen models and Bayesian estimation, researchers addressed competing risks and uncertainty quantification. The figure reflects the most sophisticated extensions of Markov methods in diabetes modeling.

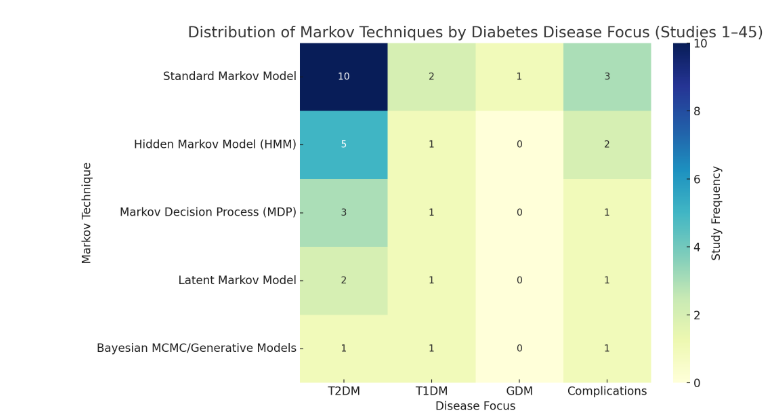
Across all 45 studies analyzed, there was a consistent and universal application of Markov-based methods, highlighting their central role in modeling chronic diseases—especially diabetes and its various subtypes and complications. The 100% usage rate observed in each group of studies (as illustrated in the figures) clearly reflects the dominance and widespread trust in these methods for handling complex, longitudinal healthcare data and disease progression modeling. A diverse range of Markov model variants were employed across these studies. Basic Markov chain models (including 2-state and 3-state models) were frequently used to represent disease stages such as glycemic control progression, diabetes onset, and related complications. Several studies adopted Hidden Markov Models (HMMs) and their extensions, such as Layered HMMs and Latent Markov Models, to capture unobserved disease states and transitions influenced by clinical indicators like HbA1c, BMI, and blood glucose levels.

Additionally, Markov decision processes (mdps) featured prominently in studies focused on optimizing treatment strategies, especially in diabetes management and policy formulation. Advanced techniques such as bayesian estimation with markov chain monte carlo (MCMC), reversible jump MCMC (RJMCMC), and hybrid models combining markov processes with cox/aalen survival analysis or newton divided differences were also observed. These were particularly useful in parameter estimation, mortality modeling, and handling model uncertainty. From a functional standpoint, the studies employed these models for a variety of purposes including duration estimation of disease phases, cost-effectiveness analysis, patient trajectory modeling (e.g., From gestational diabetes to T2DM), and treatment optimization. The integration of probabilistic sensitivity analysis (PSA) with Transition probability matrices (tpms) further reinforced the role of markov modeling in economic evaluations.

In summary, the analysis demonstrates a strong methodological trend favoring markov-based models, not only for their simplicity and transparency but also for their ability to accommodate extensions for complex real-world clinical scenarios. The evolution from simple markov chains to hybrid and bayesian-augmented models reflects a maturing field where model sophistication is tailored to meet specific healthcare challenges. This comparative insight underscores the continued relevance and adaptability of markov models in modern disease modeling and decision-making frameworks.

## Distribution of Markov Techniques by Diabetes Disease Focus

The heatmap provides a visual summary of how various Markov-based modeling techniques have been applied across different diabetes categories—type 2 diabetes mellitus (T2DM), type 1 diabetes mellitus (T1DM), gestational diabetes mellitus (GDM), and diabetes-related complications—based on a systematic review of 45 studies.



*Figure 10.* Image of the Distribution of Markov Techniques by Diabetes Disease Focus

The vertical axis represents distinct markov methodologies such as standard markov models, hidden markov models (hmms), markov decision processes (mdps), latent markov models, and markov chain monte carlo (MCMC) approaches. The horizontal axis categorizes the disease focus into T2DM, T1DM, GDM, and complications (e.g., cardiovascular disease, obesity, mortality). The color gradient in the heatmap reflects the frequency of each technique's application in the corresponding disease category, with darker shades indicating higher usage intensity. The visualization reveals the following:

Standard markov models and hmms are most frequently applied in T2DM studies, highlighting their suitability for modeling chronic disease progression and treatment outcomes.MCMC and bayesian markov techniques are used extensively for parameter estimation, especially in T1DM models, where physiological modeling often requires precise individualized inference. Markov decision processes (mdps) are predominantly employed in T2DM studies, highlighting their relevance in optimizing treatment strategies and evaluating long-term healthcare outcomes (e.g., Qalys). Latent markov models and layered hmms have been used in studies addressing complications or heterogeneous healthcare utilization patterns. The GDM is comparatively underrepresented, with only a few studies utilizing basic markov chains or state transition models, indicating a potential research gap. This distribution emphasizes the flexibility and growing interest in leveraging markov-based frameworks for personalized prediction, intervention evaluation, and health economics modeling in diabetes care. It also underscores the potential for extending underutilized methods such as latent markov modeling to underexplored domains such as GDM and multimorbidity management.

# Future Rearch

Future research on the application of Markov chain models to diabetes should focus on enhancing model accuracy, scalability, and clinical relevance through the integration of real-world, high-resolution data such as continuous glucose monitoring (CGM), electronic health records (EHRs), and wearable sensor outputs. There is a growing need to address data heterogeneity and missing values by combining Markov models with advanced machine learning techniques, such as deep learning and reinforcement learning, to capture nonlinear and temporal dependencies in disease progression. The incorporation of time-varying covariates and patient-specific characteristics will improve personalized risk prediction and treatment optimization. Additionally, further development of hybrid models, such as combining hidden Markov models with Bayesian inference or MDPs with simulation frameworks, can enhance decision-support systems in clinical environments. Emphasis should also be placed on conducting probabilistic sensitivity analyses to quantify the uncertainty in transition probabilities and their impact on long-term outcomes, including quality-adjusted life years (QALYs) and healthcare costs. Finally, collaborative efforts between data scientists, clinicians, and policymakers are essential to translate model insights into actionable public health strategies and real-time clinical decision-making tools, particularly in resource-limited settings where diabetes incidence continues to rise. For our ongoing research, we extend these findings by applying non-homogeneous discrete-time Markov models combined with machine learning techniques to real diabetes datasets. Our focus is on age-stratified progression modeling, simulation of patient trajectories using 2023–2024 clinical data, and the integration of Hidden Markov Models (HMMs) to address irregular or sparse patient records. This extension will allow us to validate the review findings in a practical setting and demonstrate the predictive and decision-support potential of Markov frameworks in diabetes management.

## Strengths and Limitations of the Review

This review provides a comprehensive synthesis of diverse computational and statistical approaches to diabetes modeling, with a particular emphasis on Markov-based frameworks. A major strength of the work lies in the breadth of methodologies examined, ranging from standard Markov chains and hidden Markov models to Bayesian estimation, simulation-based approaches, and hybrid methods that integrate machine learning. By consolidating these techniques into a unified narrative and supplementing the discussion with comparative tables and figures, the review enhances interpretability for both technical and clinical readers. Another important strength is its translational relevance, as the models reviewed are discussed not only from a theoretical perspective but also in terms of their practical applications in predicting complications, optimizing treatment strategies, and evaluating cost-effectiveness. The review also highlights emerging areas such as reinforcement learning, digital twin technologies, and generative modeling, thereby offering a forward-looking perspective for future research. At the same time, several limitations must be acknowledged. The studies included in the review are highly heterogeneous in terms of sample size, population characteristics, and data quality, which restricts the comparability of results and the generalizability of findings. In addition, most of the existing literature is centered on Western populations, leaving important gaps in region-specific modeling for South Asia and other developing regions. Another limitation is that many models remain insufficiently validated, often relying on simulated or small-scale datasets rather than large, diverse clinical cohorts, which raises concerns about their robustness in real-world applications. Furthermore, while the review captures a wide range of stochastic approaches, it does not fully reflect the most recent advances in deep learning architectures, such as transformer-based time-series models, which could be effectively combined with Markovian frameworks. Finally, much of the literature focuses on long-term projections, while short-term disease progression modeling using limited data remains underexplored. By acknowledging these limitations alongside its strengths, the review not only remains balanced but also identifies concrete research gaps, thereby providing a foundation for future work, including our own studies on Indian diabetes datasets using hybrid Markov–machine learning models.

# Conclusion

This review has shown that Markov chain models remain powerful tools for understanding the complex dynamics of diabetes, particularly in type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM), and gestational diabetes mellitus (GDM). By consolidating studies that employ standard Markov chains, hidden Markov models, Markov decision processes, and Bayesian extensions such as MCMC, the review highlights how these approaches have been successfully applied to disease progression, treatment optimization, complication prediction, and healthcare cost analysis. The ability of these models to represent transitions between health states provides valuable insights into screening, early intervention, and long-term outcome evaluation. At the same time, the synthesis emphasizes the growing importance of hybrid approaches that integrate Markov processes with machine learning and simulation-based strategies. These combinations enable more personalized predictions, improved handling of uncertainty, and greater clinical applicability. The review therefore not only reflects the progress made in computational diabetes research but also provides a foundation for future studies to build upon. In particular, our ongoing work with Indian patient datasets aims to extend this literature by demonstrating how hybrid Markov–machine learning models can simulate disease progression and inform region-specific clinical decisions. By bridging theoretical advances with real-world applications, this review contributes to the development of predictive tools that support precision diabetes care and evidence-based health policy.

# References

1. Cappon, G., Prendin, F., Facchinetti, A., Sparacino, G., & Del Favero, S. (2023). Individualized models for glucose prediction in type 1 diabetes: comparing black-box approaches to a physiological white-box one. *IEEE Transactions on Biomedical Engineering*, 70(11), 3105–3115.
2. Alkhateeb, H., El Fathi, A., Ghanbari, M., & Haidar, A. (2021). Modelling glucose dynamics during moderate exercise in individuals with type 1 diabetes. *PLOS ONE*, 16(3), e0248280.
3. Zeng, F., & Wang, Y. (2016, July). Dynamic model with time varying delay for type 1 diabetes mellitus identified by using expectation maximization algorithm. In *2016 35th Chinese Control Conference (CCC)* (pp. 9376–9381). IEEE.
4. Haidar, A., Wilinska, M. E., Graveston, J. A., & Hovorka, R. (2013). Stochastic virtual population of subjects with type 1 diabetes for the assessment of closed-loop glucose controllers. *IEEE Transactions on Biomedical Engineering*, 60(12), 3524–3533.
5. Ewings, S. M., Sahu, S. K., Valletta, J. J., Byrne, C. D., & Chipperfield, A. J. (2015). A Bayesian network for modelling blood glucose concentration and exercise in type 1diabetes. *Statistical Methods in Medical Research*, 24(3), 342–372.
6. Luo, S. L., Guo, W. D., Zhang, J., Guo, F., Chen, Y. Y., Han, L. F., & Chen, S. J. (2011). Prediction Technology of Type Diabetes Based on Markov Chain. *Transactions of Beijing Institute of Technology*, (12), 1414–1418.
7. Neumann, A., Schwarz, P., & Lindholm, L. (2011). Estimating the cost-effectiveness of lifestyle intervention programmes to prevent diabetes based on an example from Germany: Markov modelling. *Cost Effectiveness and Resource Allocation*, 9, 1–13.
8. Chilcott, J. B., Whitby, S. M., & Moore, R. (2001, August). Clinical impact and health economic consequences of posttransplant type 2 diabetes mellitus. In *Transplantation Proceedings* (Vol. 33, No. 5A Suppl, pp. 32S–39S).
9. Lin, M. Y., et al. (2021). Data analysis of the risks of type 2 diabetes mellitus complications before death using a data-driven modelling approach. *Information*, 12(8), 326.
10. Zhu, C. R., Ni, Z. Z., & Sun, K. H. (2005). A Markov model to assess long-term effects of rosiglitazone natrium on the treatment of type 2 diabetes. *Journal of Hygiene Research*, 34(3), 355–358.
11. Santoso, L., & Mareels, I. M. (2001, December). Markovian framework for diabetes control. In *Proceedings of the 40th IEEE Conference on Decision and Control* (Vol. 2, pp. 1705–1710).
12. Zhou, H., et al. (2005). A computer simulation model of diabetes progression, quality of life, and cost. *Diabetes Care*, 28(12), 2856–2863.
13. Xiao, H., Gao, J., Vu, L., & Turaga, D. S. (2017). Learning temporal state of diabetes patients via combining behavioral and demographic data. In *Proceedings of the 23rd ACM SIGKDD Conference* (pp. 2081–2089).
14. Mueller, L., et al. (2020). Application of machine learning models to evaluate hypoglycemia risk in type 2 diabetes. *Diabetes Therapy*, 11, 681–699.
15. Kuo, H. S., et al. (1999). A Markov chain model to assess the efficacy of screening for NIDDM. *International Journal of Epidemiology*, 28(2), 233–240.
16. Aliyari, R., et al. (2020). Multistate models to predict development of late complications of type 2 diabetes. *Diabetes, Metabolic Syndrome and Obesity*, 1863–1872.
17. Lovblom, L. E., et al. (2024). Modeling multiple correlated end-organ disease trajectories: A tutorial. *Statistics in Medicine*, 43(5), 1048–1082.
18. Li, J., et al. (2021). Decision models in type 2 diabetes mellitus: a systematic review. *Acta Diabetologica*, 58(11), 1451–1469.
19. Amritha, Y. D., & Dipta, I. M. Y. (2024, July). Multistate Model with Hidden Markov for Chronic Disease. In *2024 ICSCC* (pp. 208–211). IEEE.
20. Omana, J., & Moorthi, M. (2022). Predictive analysis of diabetes with ML techniques. *Wireless Personal Communications*, 127(1), 465–478.
21. Perveen, S., et al. (2020). Handling irregularly sampled longitudinal data and prognostic modeling of diabetes using machine learning. *IEEE Access*, 8, 21875–21885.
22. Vasudevan, S. (2023). Some prediction models in the study of diabetic retinopathy in Type II Diabetes Mellitus. *Various Statistical Models Approach*.
23. Varshney, M. K., et al. (2020). Estimation of transition probabilities for diabetic patients using hidden Markov model. *International Journal of System Assurance Engineering and Management*, 11(Suppl 2), 329–334.
24. Eghbali-Zarch, M., et al. (2019). A Markov decision process for modeling adverse drug reactions. *Journal of Engineering in Medicine*, 233(8), 793–811.
25. Andersen, K. E., et al. (2004). Bayesian Model Discrimination for Glucose-Insulin Homeostasis. *Poster, Aalborg University*. Retrieved from <https://vbn.aau.dk/files/13127287/Poster05a.pdf>
26. Srikanth, P. (2015). Using Markov chains to predict the natural progression of diabetic retinopathy. *International Journal of Ophthalmology*, 8(1), 132.
27. Al Ali, H., et al. (2022). Examining type 1 diabetes mathematical models using experimental data. *International Journal of Environmental Research and Public Health*, 19(2), 737.
28. Govan, L., et al. (2015). How do diabetes models measure up? A review of diabetes economic models and ADA guidelines. *Journal of Health Economics and Outcomes Research*, 3(2), 132.
29. Tomar, R. H., et al. (1998). Disease progression and cost of insulin dependent diabetes mellitus. *Journal of the Society for Health Systems*, 5(4), 24–37.
30. Begun, A., et al. (2016). Disease progression in diabetic foot patients using a multistate Markov chain model. *PLOS ONE*, 11(1), e0147533.
31. Perveen, S., et al. (2020). A hybrid approach for modeling type 2 diabetes mellitus progression. *Frontiers in Genetics*, 10, 1076.
32. Meng, F., et al. (2020). Evaluating glycemic control strategies via Markov decision process. *Acta Diabetologica*, 57, 827–834.
33. Cho, S. W., et al. (2019). Estimating lifetime duration of diabetes in Korea using a Markov model. *Journal of Korean Medical Science*, 34(Suppl 1).
34. Narendra, S. (2023). Two Hidden Layer’s Markov Model for Studying Type 2 Diabetes. *Preprint*. DOI: 10.21203/rs.3.rs-3351123/v1.
35. Goel, K., et al. (2018). Multistate Markov model for disease progression of type 2 diabetes. *Journal of Nephropharmacology*, 8(1), e04.
36. Cappon, G., et al. (2019, July). Bayesian framework to identify type 1 diabetes physiological models. In *2019 IEEE EMBC* (pp. 6914–6917). IEEE.
37. Grassi, A., et al. (2010). Markov chain probability model of glucose tolerance in postgestational diabetes. *MEDINFO 2010*, 1155–1159.
38. Wang, X., et al. (2022). Inferring progression of complications in type 2 diabetes using generative model. *BMC Medical Informatics and Decision Making*, 22(1), 174.
39. Santoso, L., & Mareels, I. M. Y. (2001, October). Markov modelling and adaptive control in T1DM. In *2001 IEEE EMBC* (Vol. 2, pp. 2068–2071).
40. Zhang, Y., et al. (2019). Probabilistic sensitivity analysis on Markov models for type 2diabetes. *Health Care Management Science*, 22, 34–52.
41. Rosato, R., et al. (2007). Cardiovascular mortality in type 2 diabetes using Markov chains. *Journal of Evaluation in Clinical Practice*, 13(3), 422–428.
42. Malim, M. R., et al. (2019). A Markov Chain Model for Diabetes Mellitus Patients. In *iCMS2017 Proceedings* (pp. 483–488). Springer Singapore.
43. Prevost, T. C., et al. (1999). Markov chain models for cataract progression in diabetic adults. *Journal of Epidemiology and Biostatistics*, 4(4), 337–344.
44. Zuanetti, D. A., & Milan, L. A. (2017). A generalized mixture model applied to diabetes incidence data. *Biometrical Journal*, 59(4), 826–842.
45. Gil, J., et al. (2019). Uncontrolled diabetes and health care utilisation: a bivariate latent Markov model approach. *Health Economics*, 28(11), 1262–1276.
46. American Diabetes Association. (2021). Standards of medical care in diabetes—2021 abridged for primary care providers. *Clinical Diabetes*, 39(1), 14–43.
47. Battelino, T., et al. (2019). Clinical targets for CGM interpretation: international consensus. *Diabetes Care*, 42(8), 1593–1603.
48. Chen, L., et al. (2015). Effect of lifestyle intervention in patients with type 2 diabetes: a meta-analysis. *Metabolism*, 64(2), 338–347.
49. Dankwa-Mullan, I., et al. (2019). Transforming diabetes care through artificial intelligence. *Population Health Management*, 22(3), 229–242.
50. Mahajan, A., et al. (2018). Fine-mapping type 2 diabetes loci using high-density imputation. *Nature Genetics*, 50(11), 1505–1513.
51. Saeedi, P., et al. (2019). Global and regional diabetes prevalence estimates for 2019. *Diabetes Research and Clinical Practice*, 157, 107843.
52. Zeevi, D., et al. (2015). Personalized nutrition by prediction of glycemic responses. *Cell*, 163(5), 1079–1094.
53. Makroum, M. A., et al. (2022). Machine learning and smart devices for diabetes management: Systematic review. *Sensors*, 22(5), 1843.
54. Contreras, I., & Vehi, J. (2018). AI for diabetes management and decision support.
55. *Journal of Medical Internet Research*, 20(5), e10775.
56. Zhu, T., et al. (2020). Deep learning for diabetes: a systematic review. *IEEE Journal of Biomedical and Health Informatics*, 25(7), 2744–2757.
57. Herrero, P., et al. (2017). A coordinated insulin-glucagon delivery control strategy in type 1 diabetes. *Computer Methods in Biomechanics and Biomedical Engineering*, 20(13), 1474–1482.
58. Man, C. D., et al. (2014). The UVA/PADOVA type 1 diabetes simulator: new features. *Journal of Diabetes Science and Technology*, 8(1), 26–34.
59. Toffanin, C., et al. (2013). Artificial pancreas: model predictive control design from clinical experience. *Journal of Diabetes Science and Technology*, 7(6).
60. Razavian, N., et al. (2015). Population-level prediction of type 2 diabetes from claims data. *Big Data*, 3(4), 277–287.
61. Arumugam, R., & Rajathi, M. (2019). Applications of manpower levels for business with various recruitment rates in the system through stochastic models. International Journal of Recent Technology and Engineering, 8(2S4), 719–722.