Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Business Analytics and Quantitative Marketing

# Evaluating Treatment Outcomes for Bovine Mastitis: A Comparative Analysis of Traditional Multi-State Models and the Innovative MSRIST Approach

Martino Martini (580781)

| | |
|---|---|
| Supervisor: | Prof. dr. Patrick Groenen |
| Second assessor: | Prod. dr. Richard Paap |
| Date final version: | November 16, 2023 |

**Abstract**

The research indicates that although the MSRIST algorithm demonstrates superior performance over the Cox Propotional Hazard Frailty Approach and Cox Proportional Hazard models in predicting state transitions, the distinctions in their predictive behaviors are subtle and become more apparent under detailed examination. Cows treated with antibiotics show an increased propensity to enter the "Death" state, a trend that can be attributed to their use in critical cases. On the other hand, Quick-Extra is emerging as a promising natural alternative for early intervention, ensuring that milk integrity remains intact and posing no risk to consumer health, thereby extending the duration of cows' presence on the farm. The need for a structured trial to refine these observations is evident. In addition, the research underscores the critical influence of unobserved heterogeneity on state transition outcomes and the complexity of directly comparing the two treatments due to their different timing of administration. While preliminary results from our dataset support our findings, further research is needed to draw definitive conclusions.

# Contents

# 1 Introduction

The dairy sector has a crucial impact on the global agricultural industry as it offers essential nutritional products and significantly affects economies worldwide. A major issue that is becoming increasingly important in this sector is the treatment of mastitis, an inflammatory condition of the mammary gland in dairy cows mainly caused by bacterial infections. Mastitis is regarded as the most prevalent disease responsible for economic losses in dairy industries causing decreased milk yield and substandard milk quality (Gomes & Henriques, 2016). Focusing on the financial implications of mastitis, it generally costs a farmer approximately $150 per cow per year due to bovine mastitis (Cheng & Han, 2020). This cost, primarily due to decreased milk output and culling, amounts to 11% to 18% of a cow's annual gross profit (Hogeveen et al., 2019). Notably, 70% of these costs are a result of diminished milk yield due to mammary tissue damage (Zhao & Lacasse, 2008). Mastitis is mainly classified into two classes based on the degree of inflammation, namely clinical and sub-clinical mastitis. Visible signs, such as a reddened, swollen udder and fever in a dairy cow, indicate the presence of clinical bovine mastitis. The cow's milk often appears diluted and may contain clumps and particles (Khan & Khan, 2006). Depending on the intensity of the inflammation, clinical mastitis is classified into per-acute, acute, and sub-acute stages (Kibebew, 2017). In severe cases, clinical mastitis may lead to fatality (Ruet et al., 2001). On the other hand, sub-clinical mastitis presents no visible signs in either the udder or milk but results in decreased milk yield and an elevated somatic cell count (SCC) (Abebe et al., 2016). While measuring the impact of sub-clinical mastitis is challenging, it is generally believed to cause more economic strain on the herd than its clinical counterpart. This is evident from studies conducted by Romero et al. (2018), N. Sharma et al. (2011), and Zhao and Lacasse (2008) which emphasize its economic consequences.

Research in Sharma (2011) revealed that 62.6% of the cows studied experienced mastitis, with a 95% confidence interval ranging from 58.3% to 66.7%. The statistics breakdown further revealed that sub-clinical mastitis was more widespread, affecting 59.2% of the cows. Sub-clinical mastitis is detectable through testing but not visible through observation. Only 3.4% of the cows exhibited clinical mastitis, which was determined through visible symptoms such as swelling and redness. There was a noticeable contrast between these animals and the rest.

The main strategy to treat mastitis is by the use of antibiotics, such as penicillin, ampicillin, tetracyclin, gentamycin, etc., which can be given by intra-mammary infusion, intramuscular or intravenous injections (Hossain et al., 2017). The overuse and misuse of antibiotics in the

treatment of bovine mastitis has caused some problems in the dairy industry (Cheng & Han, 2020). Furthermore, the presence of antibiotic residues in milk also raises notable concerns. In general, milk obtained during antibiotic treatment, followed by a waiting period, must be discarded since it cannot be consumed due to the risk of allergies and drug resistance caused by antibiotic residues (Gomes & Henriques, 2016). Heavy penalties are charged for antibiotic residues in milk. However, some antibiotic residues remain in the animal body beyond their recommended discard times. The cost of treatment depends on the loss from discarded milk rather than the cost of drugs. Additionally, while antibiotics can eliminate infection, they do not provide direct protection to the mammary gland from irreparable damage (Cheng & Han, 2020). Farms constantly endure reduced lifetime milk productivity, resulting in losses (Zhao & Lacasse, 2008).

In response to these challenges, Animal Health Vision (AHV) has pioneered a novel therapeutic strategy termed "Quorum Quenching" (QQ). Quorum Quenching hinders bacterial biofilm formation by disrupting their communication capabilities (X. Zhu et al., 2022). Studies such as the one conducted by Abranches et al. (2011) have identified certain plant-based compounds that could offer a promising alternative to conventional antibiotics. Given the evolving treatment methods and the complexities of understanding and predicting health outcomes, advanced analytical tools are essential. Time-to-event outcomes prevalent in fields like medicine, epidemiology, and environmental health offer insights into the occurrence and timing of pivotal events, such as disease or death (Le-Rademacher et al., 2022b). The Cox Proportional Hazards model is commonly used for analyzing time-to-event data and is particularly suited for singular events, as noted by (Cox, 1972). In cases where multifaceted events or factors are involved, conventional methods may prove inadequate. This emphasizes the necessity for sophisticated analytical techniques in complex scenarios.

This is precisely where Multi State Models (MSMs) distinguish themselves as a potent and adaptable instrument. They can shed light on treatment impact on intermediate events and detail the different trajectories cows may traverse in these complex contexts. This study uses multiple MSMs to compare the impact of the Quick-Extra (QE) treatment from AHV on (clinical) mastitis against that of antibiotics. MSMs offer insights into the effects of treatment on intermediate events and clarify potential pathways that cows may follow in complicated scenarios (P. K. Andersen & Keiding, 2002). We also use the MSM to investigate the transition nature of (clinical) mastitis by examining transition probabilities and how these are influenced by different covariates (Putter et al., 2007b).

Estimating transition probabilities can be achieved through various methods. In the absence of covariates, Aalen and Johansen (1978) utilized counting process methods (Khan & Khan, 2006). When covariates are present, it is essential to factor in their impact on transition intensities. To do so, commonly used regression models include Cox Proportional Hazards (CPH) regression (Lakew et al., 2019).

Recently, machine learning methods have emerged as prominent tools (Hothorn et al., 2004; Ishwaran & Kogalur, 2010a; Schmid et al., 2016). These methods focus on making predictions that work well for new, unseen data (Van Belle et al., 2008). Our research employs a CPH estimator to implement a conventional MSMs, both with and without the frailty approach. Additionally, we present the Multi State Recursively Imputed Survival Trees (MSRIST) that integrates machine learning techniques. Our objective is to assess the performance of MSRIST in relation to traditional MSM techniques. With this context in mind, we present the following research question:

"How do Quick-Extra treatment outcomes for mastitis compare to antibiotics when evaluated using MSMs and MSRIST, and can machine learning enhance the accuracy of conventional MSMs?"

This research compares conventional MSMs with the newly proposed MSRIST and evaluates them using Concordance index and Integrated Brier Score. We evaluate treatment outcomes for mastitis using actual clinical data, striving for an in-depth understanding of treatment efficacy. In the Literature Section 2, we begin with an overview of mastitis treatments, transition to a discussion of MSMs, touch on estimation techniques, and conclude with an exploration of the integration of machine learning with MSMs. Datasets employed are detailed in Section 3. The Methods section, starting at Section 4, introduces MSM basics and advances through key models like the Cox Proportional Hazard Model with and without the Frailty Approach and the Random Survival Forest. Our Results, presented in Section 5, include comparisons of model performance and in-depth analysis of treatment effects. We conclude and discuss implications in Section 6.

# 2 Literature Review

In this chapter, we will examine the literature on treatments for mastitis, with a focus on the potential of Quorum Quenching (QQ). Afterwards we review the academic background on Multi State Models. Additionally, the integration of machine learning and Multi State Models.

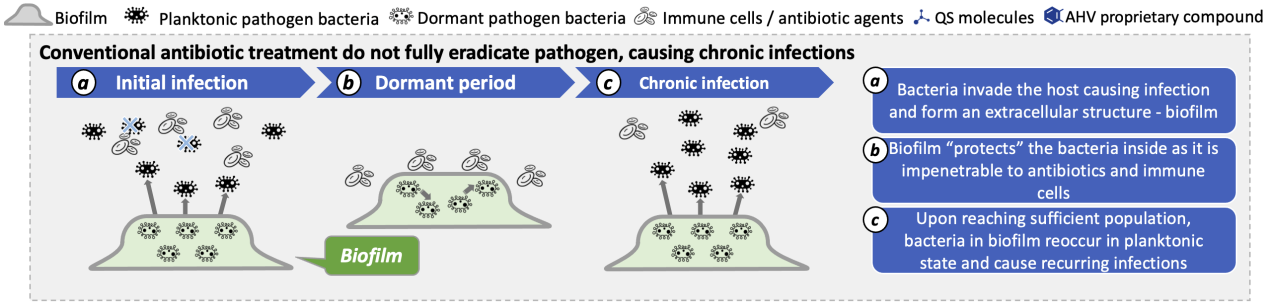
## 2.1 Alternative Treatments for Mastitis

The traditional method of treating mastitis has primarily depended on antibiotics. Nevertheless, the literature reveals a growing interest in alternative therapies and preventive measures. Natural remedies have demonstrated potential efficacy in the treatment of mastitis (Down et al., 2019). Additionally, probiotics are being examined as a possible therapeutic measure (Markowiak & Śliżewska, 2018). Homeopathic interventions are also under consideration, with research such as Tomanić et al. (2023) exploring their potential. However, the effectiveness of these treatments remains variable. Preventive options include teat sealants and immunotherapies, with a study by Sharun et al. (2021) underscoring their role in preventing the onset of mastitis. An especially significant challenge in treating mastitis is the formation of biofilms by the pathogens responsible. Structured bacterial colonies can reduce the efficacy of conventional treatment approaches. Therefore, researchers are exploring methods to identify the quorum sensing (QS) system that fosters bacterial communication and biofilm formation. The aspiration is to impede biofilm evolution by interrupting the QS system, thus enhancing bacterial sensitivity to treatments (Davies et al., 2017). The relationship between biofilm and QS shows the promise of plant-based methods from AHV. These natural compounds could break down bacterial clusters or prevent them from forming. An illustration of the effect of an antibiotic treatment and AHV treatment is followed in Figure 1.

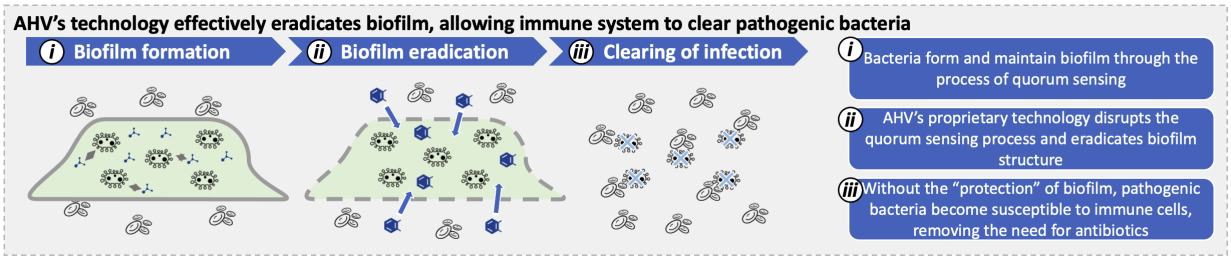
The growing concerns surrounding antibiotic resistance and the challenges posed by biofilms emphasize the need for alternative treatments. Nature-inspired therapies and interventions could hold the key to the future of mastitis treatment (Gupta et al., 2019).

## 2.2 Multi State Models

Multi State Models (MSMs) are strong frameworks for modeling individuals' experiences over time, as described in literature sources like P. K. Andersen et al. (2012) and Hougaard (2000). They are useful in a range of fields, from tracking disease progression to enhancing system reliability in engineering. The use of MSMs in chronic disease monitoring is apparent when




(a) Antibiotic treatment to bacteria




(b) AHV treatment to bacteria

Figure 1: Antibiotic- and AHV treatment on bacteria

examining the transition from pre-diabetes to type II diabetes and post-orthopedic surgery rehabilitation processes (K. Andersen & Keiding, 2002). These models aid clinicians in making informed decisions by identifying the significant covariates that affect transition probabilities (Muthén & Asparouhov, n.d.). MSMs have applications beyond healthcare, specifically in engineering and systems analysis. They are useful when studying machinery and components that are prone to wear, degradation, and malfunction. Engineers can use MSMs to predict failures, schedule maintenance, and optimize system performance by assessing transitions between operational states (Limnios & Oprisan, 2001; Yingkui & Jing, 2011). The field of stochastic processes includes a variety of models, such as the Continuous-time Markov Model (CTMM), that showcase different systems and transitions (Wan et al., 2017). Nevertheless, these models possess the Markov property, indicating that state transitions rely solely on the present state (Markov, 1954). For situations that require a wider range of transition dynamics, Semi-Markov models are the preferred choice due to their inherent flexibility (Jackson, 2011; Pérez-Ocón et al., n.d.). Conversely, non-Markov models, which consider historical data for transitions, provide a more complex context for understanding state dynamics (Bobbio et al., 1998; van Kampen, 1998). However, the computational costs and data requirements can be challenging due to their complexity (Isensee et al., 2023). Additionally, non-Markov models, despite their flexibility,

may not always provide superior predictive accuracy, as stated by Tapak et al. (2018).

## 2.3 Estimation Methods Background for Multi State Models

The accurate estimation of transition probabilities in Multi State Models (MSMs) is a pivotal challenge. Such estimation is crucial for discerning the system's dynamics and predicting its behavior (P. K. Andersen & Keiding, 2002). Literature posits two predominant methods for this estimation: the Aalen-Johansen technique and the Cox Proportional Hazard (CPH) model (Le-Rademacher et al., 2022b). A distinguishing feature of the Aalen-Johansen method is its reliance on sojourn times and its omission of covariates (Aalen et al., 2008; Aalen & Johansen, 1978). Given our focus on contrasting treatments, which inherently function as covariates, this highlights the necessity of incorporating covariates in method comparison. Hence, we contend that the CPH model, with its ability to effectively integrate covariates, is more applicable (Cox, 1972). Nonetheless, addressing the issue of unobserved variables remains paramount, as elaborated below.

Emerging in the second half of the 20th century, the frailty approach facilitates the management of unobserved variance in data (Vaupel et al., 1979). The underpinnings of this model lie in latent variables, or "unobservable heterogeneities", that significantly influence subjects' susceptibilities and responses to events or conditions (Springer, 2008). A case in point is studies of bovine mastitis, where frailty models elucidate intrinsic differences in susceptibility and response to the disease on a per-cow basis or management practices. Its versatility is evident in several areas of biomedical research (Aalen et al., 2008; Duchateau & Janssen, 2008; Wiener & Tilly, 2002). For example, it is useful in analyzing recurrent event data, such as iterative hospital admissions (Hougaard, 1999), and in epidemiologic studies, where it illuminates variations in disease susceptibility due to unseen genetic factors or other latent variables (Rondeau et al., 2003). Comparison between frailty and non-frailty models reveals latent differences that may influence outcomes, underscoring the need to carefully examine latent determinants and guide subsequent research directions (Putter et al., 2007a).

## 2.4 Machine Learning in Multi State Models

The article by Tapak et al. (2018) underscores the integration of machine learning in Multi State Model analysis. The use of machine learning techniques has significantly expanded in fields that demand time-to-event data analysis (Tapak et al., 2018). Data-driven ensemble methods, a category within machine learning, have attracted the attention of numerous researchers (Hothorn

et al., 2004; Ishwaran & Kogalur, 2010b; Schmid et al., 2016). These methods stand out for their capacity to select relevant covariates and predict survival data even when censoring is present. To clarify, censoring occurs when complete information about the timing of an event is unavailable because data collection was terminated before the event took place or is still ongoing during analysis. For instance, if the observation period concludes after 12 months, and a patient has not yet reached the endpoint of interest, this is an example of censoring, wherein the exact timing of the event remains unknown due to the study's limited duration. The main objective of data-driven ensemble techniques is to construct a model that can accurately predict future unobserved data instances (Van Belle et al., 2008). Such methods are particularly valuable in situations that require precise prediction of results. Random forests (RF) is a prominent approach within the ensemble learning domain, which is widely used in machine learning and data mining. One methodology for examining survival data is the random survival forests (RSF) method (Ishwaran et al., 2008). RSF overcomes challenges associated with conventional methods by integrating concepts from adaptive nearest neighbors and bagging (Ishwaran et al., 2008). Additionally, it offers a unique ability to select and prioritize variables through the use of variable importance assessments. While recognizing certain constraints of the RSF model proposed by Ishwaran and Kogalur (2010a), particularly its requirement for a minimum number of observed failure events in terminal nodes, which makes it challenging to incorporate censored observations, R. Zhu and Kosorok (2012) sought to improve the model's performance by implementing optimizations. Their method, named recursively imputed survival trees (RIST), is an exceptional nonparametric imputation approach. The process consists of iteratively updating censored observations to correspond with current model-based conditional failure times and fitting the model again with the refined data. By repeatedly following this procedure, the RIST approach seamlessly integrates the conditional failure times of censored observations into the model. The advantages of combining ML with MSMs include lessened prediction error and enhanced model accuracy (R. Zhu & Kosorok, 2012). Currently, there is limited literature on this topic. However, Tapak et al. (2018) proposes a method for combining Multi State with randomly imputed survival tree, referred to as Multi State RIST (MSRIST). The research indicates that MSRIST outperforms traditional Cox-Hazard MSMs in predicting outcomes. However, this applies solely to a unidirectional model without the possibility of returning to a previous state. No literature has been found on the application of MSRIST to a bi-directional model. Consequently, it would be intriguing to examine whether the identical outcome holds for bi-directional models.

# 3 Data

The data preparation phase encompasses activities essential for establishing the finalized dataset for analysis. Several pre-processing actions were undertaken prior to model prediction. However, before delving into the data, it is crucial to first understand the background from the literature on the subject. This Section will provide a context for interpreting the datasets we have obtained.

## 3.1 Academic Background

When a cow is producing milk, it is milked daily by a milk robot which tracks the milk yield and calculates the Somatic Cell Count (SCC). According to the literature, such as the study by N. Sharma et al. (2011), an SCC above 250,000 cells/ml suggests that the cow may have clinical or subclinical mastitis. Various factors can influence a cow's SCC. While sickness is a recognized cause, other contributing factors include the physical structure of the udder (Waller et al., 2014). The conformation of a cow's udder can directly impact its vulnerability to infection and overall udder health. Furthermore, research as Król et al. (2013) highlights that older cows are at a higher risk of mastitis. Additionally, individual variations in the immune system of cows can also impact their vulnerability. As these variables cannot be directly observed using a milking machine, they should be classified as unobserved.

## 3.2 Milk Robot Dataset

After considering the literature, we turn our attention to the collected datasets. There are two major datasets at our disposal. One dataset tracks daily milking robot data from six Dutch farmers, while the other dataset is a combination of Dutch Cow Register (CRV) and government data, known as the Milk Production Records (MPRs). The daily milking robot dataset, sourced from six individual farms for one lactation phase in 2021, is particularly useful for daily analysis. However, this study's limitation is its focus on cows from specific farms and lack of data on cow mortality. Conversely, the CRV and government dataset (CRV, 2023) offers a more comprehensive view, including insights into cow mortality and making it suitable for Multi State Models studies. Nevertheless, the daily data remains invaluable as it offers a glimpse into the progression of mastitis and the effects of various treatments. Our analysis commences with the daily data, characterized by 173 cow antibiotic treatments and 36 Quick-Extra treatments from AHV. Each cow in this study is 100% Holstein-Friesian. An overview of

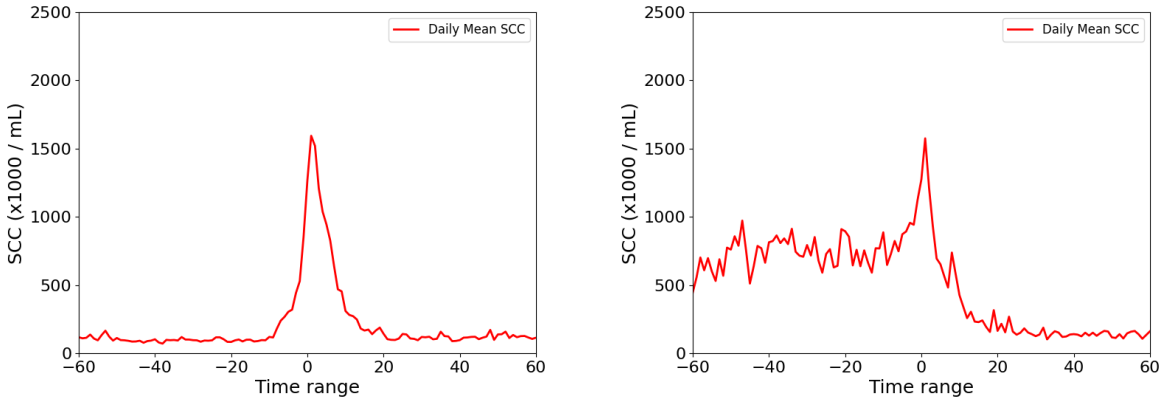
the five farms participating is shown in table 1.

Table 1: Farm details

| Farm | A | B | C | D | E |
|---|---|---|---|---|---|
| Herd size | ±250 | ±185 | ±100 | ±180 | ±180 |
| Grazing | Yes | No | Yes | No | Yes |
| Bedding | Sawdust | Dried manure | Dried manure | Flax-lime | Box Powder |
| Location | South-Holland | South-Holland | North-Brabant | Gelderland | North-Holland |

Before we dive into the details of mastitis progression, it's important to understand how we've tracked its development. By looking at a specific time period, we can shed light on the different stages a cow goes through, both before and after a case of mastitis. To gain a clearer understanding of mastitis progression over time, we observed a 120-day period, including 60 days before and after the identified mastitis case. We designated the day of treatment as day 0 and presented this timeframe in a summarized form across various phases in Figure 2. It is noteworthy that our selection of day 0 is based on treatment initiation, which typically coincides with the peak SCC observed in most cases. Figure 2a depicts a healthy cow that develops mastitis, is treated on day 0, and subsequently recovers to a healthy state. The SCC remains below 250,000 cells/ml before and after the mastitis occurrence. Figure 2b illustrates a cow that starts with an SCC above 250,000 cells/ml, indicating subclinical mastitis. After developing clinical mastitis, she is treated on day 0 and ultimately returns to a healthy state with an SCC below 250,000 cells/ml. Figure 2c illustrates a cow that initially has subclinical mastitis (SCC above 250,000 cells/ml). Despite receiving treatment on day 0 following the onset of clinical mastitis, she does not fully recover and continues to exhibit subclinical mastitis, as evidenced by an SCC that remains above 250,000 cells/ml. Finally, Figure 2d depicts a cow that, although healthy at first, develops mastitis but fails to fully recover after treatment. As a result, the cow displays indications of subclinical mastitis, with a SCC that consistently exceeds 250,000 cells/ml. It is crucial to note that even after treatment, some cows retain an SCC above the 250,000 cells/ml limit, indicating a weakened immune response. This sustained elevation is common in cases such as sub-clinical mastitis, where the cow remains in an extended state of illness rather than promptly recovering their health. Exploring the factors that influence these transitions provides valuable insights into the dynamics of mastitis development and healing.

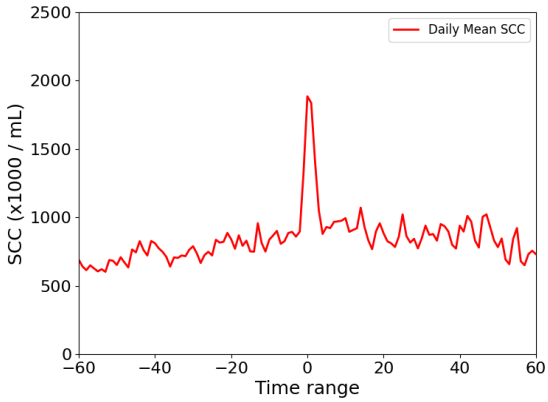
Figure 3 provides insight into the effects of different treatments on daily SCC levels. Ob-



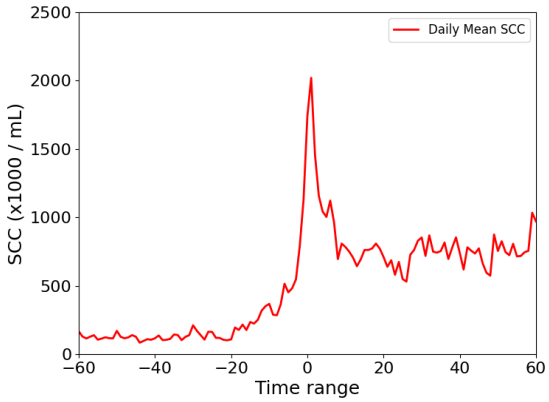

(a) Category I




(b) Category II




(c) Category III




(d) Category IV

Figure 2: Four Distinct Patterns of Somatic Cell Count Progression, where day 0 indicates mastitis treatment

servations were conducted when mastitis was identified on day zero, after which the cows were treated with either antibiotics or Quick-Extra treatment. The terms "cat high" indicate an SCC higher than 250,000 cells/ml, while "low" indicates levels below this threshold. Our findings suggest that antibiotic and Quick-Extra treatments produce similar results in terms of SCC levels. However, it is important to note that these findings are based on a restricted sample of 36 Quick-Extra (QE) treatments and 90 antibiotic treatments.

### 3.3 Milk Production Records Dataset

We created a full dataset by carefully combining farmers' MPRs with additional data sources. When using MSMs, it's important to consider cow mortality, which creates an absorptive state, as explained in relevant studies. For this purpose, we have merged information from both CRV

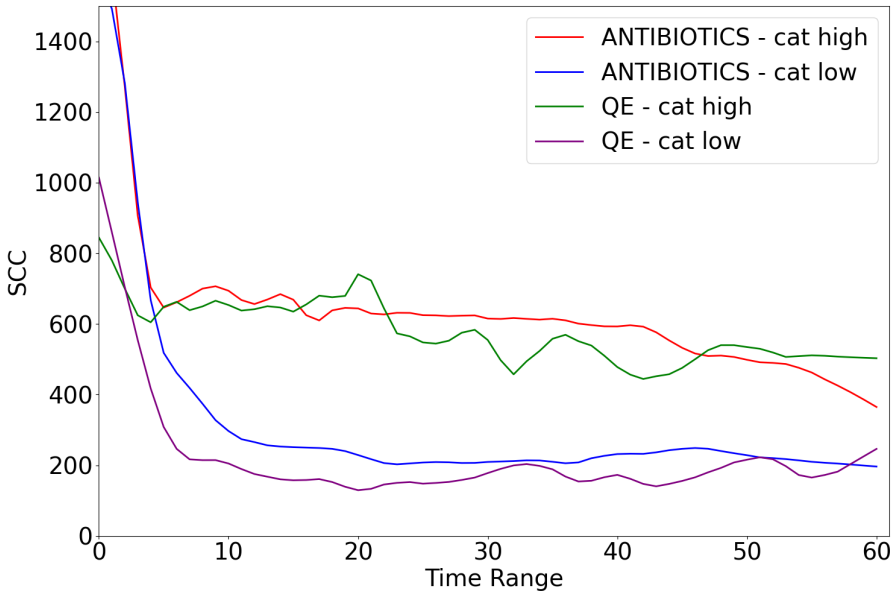



Figure 3: SCC trendlines for antibiotics- and QE treatments

and RVO sources. The resulting dataset is a fusion of the MPRs and animal movement data obtained from the RVO, enriched with detailed records of registered antibiotic treatments for clinical mastitis. We filtered these treatments and linked them to individual cows, and provide a comprehensive overview of the antibiotic treatments for (clinical) mastitis can be found in the Appendix A Table 10. Figure 4 provides a graphical illustration of the data compilation process.

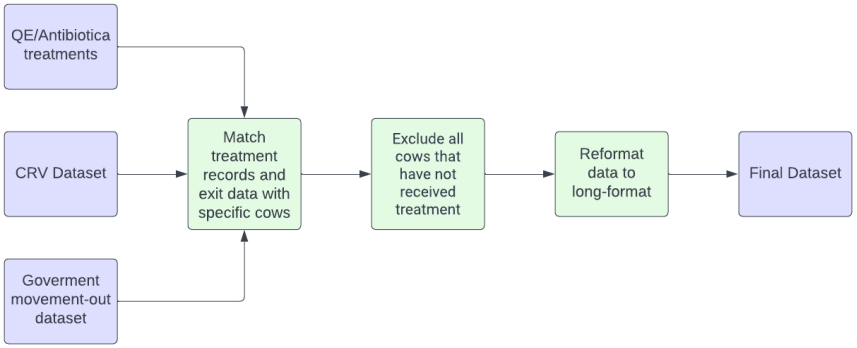



Figure 4: Overview of the steps to create the final dataset

The MPR dataset, which covers the period from 2016 to 2023, includes records for 21,983 animals with an average age of 4.94 years. This figure reflects the typical dairy farm population, where cows start producing milk at around 2 years of age and can continue to do so for up to

around 7 years before they are typically sent to a slaughterhouse. The average age in the dataset may seem low, but it reflects the productive lifespan of dairy cows, ensuring that the dataset accurately reflects the real world situation without bias towards younger or healthier animals. On average, observations are recorded at 41-day intervals for each cow. For the MSM application, we have adopted a "long" format. Table 2 illustrates this format using cow NL583454999 as an example. Each observation period begins with an initial MPR date and ends with the next MPR date, which then serves as the starting point for the next observation. All associated variables are measured relative to this start time. For example, if the SCC escalates from 100 to 400 between 2020-01-01 and 2020-02-01, the end state is classified as "Sick". The following observation will start from this state. The "Death" designation is determined by the "Movement Out" date, which indicates a cow's removal from the farm, typically to a slaughterhouse. Beyond this date, the cow's records end.

Table 2: Long-format from cow NL583454999

| Start_Time | Stop_Time | Start_State | Stop_State | Age | QE | Anti | scc | days | Event |
|---|---|---|---|---|---|---|---|---|---|
| 2020-12-24 | 2021-03-19 | Healthy | Healthy | 7 | 0 | 0 | 171 | 85 | 0 |
| 2021-03-19 | 2021-04-30 | Healthy | Sick | 8 | 0 | 0 | 87 | 42 | 1 |
| 2021-04-30 | 2021-06-13 | Sick | Healthy | 9 | 0 | 0 | 297 | 44 | 1 |
| 2021-06-13 | 2021-07-24 | Healthy | Healthy | 9 | 0 | 0 | 143 | 41 | 0 |

In addition, in Table 3 below provides a demographic overview of the animals studied, categorized based on certain conditions of treatments to which the cows were subjected. It shows the size of each group, the age range, the mean age and quarterly ranges for age, and the median stage of lactation. The lactation phase for cows refers to the period during which a cow produces milk following the birth of a calf.

Table 3: Demographic overview of animals by treatment group

| Group | No. Treated | Min Age | Max Age | Avg Age | Q25 Age | Q50 Age | Q75 Age | Median Lactation Phase |
|---|---|---|---|---|---|---|---|---|
| Antibiotica | 2.476 | 2 | 12 | 5.4 | 4.0 | 5.0 | 7.0 | 4 |
| QE | 19.834 | 2 | 14 | 4.9 | 3.0 | 5.0 | 6.0 | 3 |

### 3.4 Understanding Multi State Models: Foundations and Key Concepts

The research of Le-Rademacher et al. (2022a) forms the basis for the explanation of Multi State Models in this section. Multi State Models are frameworks that use continuous time processes to describe and model the experiences of individuals over time (P. K. Andersen et al., 2012; Hougaard, 2000). As described in Le-Rademacher et al. (2022a), Multi State Models consist of two crucial elements: the state(s) and the transition(s). "State" refers to the changing state of a subject over time, while "transition" describes the movement from one state to another in a particular direction. The direction can be unidirectional, meaning it can only move in one direction, or bidirectional, meaning it can move between two states. A state can also be either transient or terminal. If a subject can move from one state to another, the state is considered transient. In our case we have two bidirectional transitions, namely "Healthy" and "Sick". In contrast, a state is considered terminal (or absorbing) if a subject cannot move to another state, such as death. An absorbing state may be biological, such as death, or it may be of research interest, such as in a competing risk model where all states (except the initial state) are absorbing.

To gain insight into the state transitions of cows, we analyzed the transition from the initial state to the stop state. During this interval, a cow could remain healthy, become sick, or die. For an overview of these shifts, we present the transition percentages derived from our dataset in Table 4.

Table 4: Transition probabilities from Start State to Stop State

| | Stop State | | |
|---|---|---|---|
| Start State | Died (%) | Healthy (%) | Sick (%) |
| Healthy | 15.89 | 77.90 | 6.20 |
| Sick | 17.61 | 38.22 | 44.17 |

Table 5 provides a detailed examination of the effects of two primary treatments on cows: "Quick-Extra" and "antibiotics". The outcome after treatment is influenced by the initial health status of the cow, whether she starts healthy or sick. It's important to note that our observation period for this study was limited. Therefore, some cows may still have been in the "sick" state at the end of our observation and may not have had sufficient time to either fully recover or, unfortunately, passed away. This particular aspect addresses the observation of many cows remaining in the "sick" state in the Table. As an illustrative example, let's consider

cows that started in a healthy state and underwent the "antibiotic" treatment. As shown in Table 5, 85.8% of these cows remained healthy, highlighting the effectiveness of antibiotics in maintaining health. However, 10.14% transitioned to a "sick" state and there was a mortality rate of 4.06%. On the other hand, when looking at the cows treated with "Quick-Extra", those that started out healthy had a lower mortality rate of 0.70%, with 83.48% remaining healthy after treatment. Detailed percentages for various conditions and the overall effect of both treatments are shown in Table 5. There is a notable difference when looking at overall life expectancy based on treatment. Cows treated with "Quick-Extra" have a median lifespan of 260 days, indicating that they tend to stay on the farm longer. In contrast, cows treated with antibiotics have a median lifespan of only 230 days.

Table 5: Transition probabilities by treatment and initial state

| | | Stop State | | |
|---|---|---|---|---|
| Start State | Treatment | Died (%) | Healthy (%) | Sick (%) |
| Healthy | Quick-Extra | 0.70 | 83.48 | 15.83 |
| Sick | Quick-Extra | 1.49 | 33.35 | 65.16 |
| Healthy | Antibiotica | 4.06 | 85.80 | 10.14 |
| Sick | Antibiotica | 9.21 | 37.42 | 53.37 |

# 4 Methods

This section examines the current use of Multi State Modles (MSMs) in the literature and discusses the different estimation methods used to determine the transitions. A machine learning method designed for MSMs is also discussed. Additionally, we elaborate on the methodologies used to evaluate prediction performance.

## 4.1 Definitions of Symbols

In this section, we list and explain the symbols used in our study of how and when certain events occur, specifically within a framework known as the Multi-State Model (MSM). These symbols represent various mathematical elements that help us understand the timing and probability of events, such as the chance of someone surviving past a certain time, or the risk of an event happening at a certain time. We also cover symbols that describe possible changes in state over time, how often these changes occur, and factors that may influence these changes. Table 6 is an important reference to understand the terms and calculations used throughout our analysis.

Table 6: Definitions of symbols used in the MSM framework.

| **Symbol** | **Definition** |
|---|---|
| $T$ | Duration from a starting point to a terminal event. |
| $G$ | Total number of states in MSM. |
| $T = [0, \tau]$ | Interval denoting time in MSM, with $\tau$ marking the study's conclusion. |
| $A(t)$ | Intensity matrix $G \times G$, filled with hazard functions for each state transition. |
| $P_{gg'}(s, u)$ | Transition probability in MSM from state $g$ at time $s$ to $g'$ at $u$. |
| $Z$ | Covariate vector in the Cox proportional hazard model. |
| $\beta_{gg'}$ | Vector of regression coefficients for transition from $g$ to $g'$ in Cox model. |
| $\omega$ | Random effect in frailty models. |
| $P_{gig'}$ | Estimated probability of individual $i$ transitioning from state $g$ to $g'$. |
| $O_{gig'}$ | Observed status of transition for individual $i$ from state $g$. |
| $d_{ij}(t)$ | Indicator function, 1 if a transition from $i$ to $j$ is observed at $t$. |
| $N$ | Total number of individuals in the study. |

## 4.2 Foundations of Survival Analysis

Survival analysis offers insights into the distribution of time from a particular starting point, such as birth or the commencement of a treatment, leading up to a distinct terminal event, for instance, death or a relapse. Representing the duration from the onset to the event in question, the random variable $T$ is introduced. The survival function, $S(t)$, provides the probability of an individual surviving past time $t$. It is defined as

$$S(t) = P(T > t). \tag{1}$$

The cumulative distribution function of $T$ is symbolized as $F(t)$, where $F(t) = P(T \leq t)$. Then, the survival function mirrors the inverse of this cumulative distribution, illustrated by $S(t) = 1 - F(t)$. An alternative expression of the survival function is

$$S(t) = \int_t^\infty f(v)\, dv. \tag{2}$$

Here, $f(v)$ stands for the probability density function (pdf) of the random variable $T$. The integral essentially depicts the expanse beneath the curve of the pdf $f(v)$ from a specified time $t$ onwards to infinity. This conveys the probability of a person's survival past the time $t$, equivalently represented by the entire region to the right of $t$ on the pdf.

Furthermore, the hazard rate function, denoted as $h(t)$, sheds light on the probability of an individual, who has not yet confronted an event by time $t$, undergoing that event shortly after. Formally given as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T > t)}{\Delta t}. \tag{3}$$

Similarly, the cumulative hazard function is

$$H(t) = \int_0^t h(v)\, dv. \tag{4}$$

Serving as an indicator for the likelihood of an event's occurrence, for a continuously variable $T$, the interplay between survival and cumulative hazard is encapsulated by

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(v)\, dv\right). \tag{5}$$

Building on this foundation, we transition to discussing MSMs tailored for mastitis. An MSM can be characterized by a stochastic process $X(t)$ where $t$ ranges over interval $T$, and it has a limited set of states $S = \{g_1, g_2, ..., g_G\}$, where $G$ represents the total number of states. The interval $T = [0, \tau]$ denotes time, where $\tau$ marks the study's conclusion. The variable $t$ measures

the time since a pivotal event, such as the initial diagnosis of mastitis. Within the MSM framework, the emphasis is not just on predicting events but also on identifying risk factors crucial for each transition, particularly from state $g$ to state $g'$.

### 4.3 Transition Intensities and Probabilities in Multi State Models

The instantaneous risk of transition between states in a system is quantified by the hazard function. At any given time $t$, the likelihood of the system moving from state $g$ to state $g'$ is defined by the hazard function as

$$\alpha_{gg'}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(X(t + \Delta t) = g' | X(t) = g)}{\Delta t}. \tag{6}$$

However, not all transitions are feasible. Due to the inherent structure of some systems, transitions between certain states might be impossible, leading to $\alpha_{gg'}(t) \equiv 0$. This essentially indicates a zero transition intensity between such states.

To encapsulate these transition intensities across all states, we construct the intensity matrix $A(t)$. This matrix, which spans $G \times G$ dimensions (where $G$ denotes the number of possible states), is populated with hazard functions for states that differ. Specifically, each off-diagonal element is given by $\alpha_{gg'}(t)$ when $g \neq g'$. The diagonal entries, on the other hand, denote the rate at which a state exits itself and are expressed as

$$\alpha_{gg}(t) = -\sum_{g \neq g'} \alpha_{gg'}(t). \tag{7}$$

The negative sign is a reflection of this exit rate, summing all transition intensities out of state $g$ to every other state.

With the framework of transition intensities established, our primary interest centers on the transition probability $P_{gg'}(s, u)$. This probability, crucially, describes the chance that if our system was in state $g$ at an initial time $s$, it would transition to state $g'$ by a later time $u$. Mathematically, this transition is represented as

$$P_{gg'}(s, u) = \mathbb{P}(X(u) = g' | X(s) = g). \tag{8}$$

Through this mechanism, the interplay of instantaneous risks, system structure, and time dynamics allow us to predict state transitions with precision.

### 4.4 Time Scale in Multi State Models

In Multi State Models, the definition of time $t$ for hazard functions, particularly those quantifying transitions between states $g$ and $g'$, can be approached in two distinct manners:

- **Clock Forward:** Here, time $t$ signifies the duration since the patient's entry into the initial state. It starts at 0 upon entry and progresses continuously forward.
- **Clock Reset:** Upon every entry into a new state, time $t$ is reset to 0. Thus, it represents the duration exclusively within the present state.

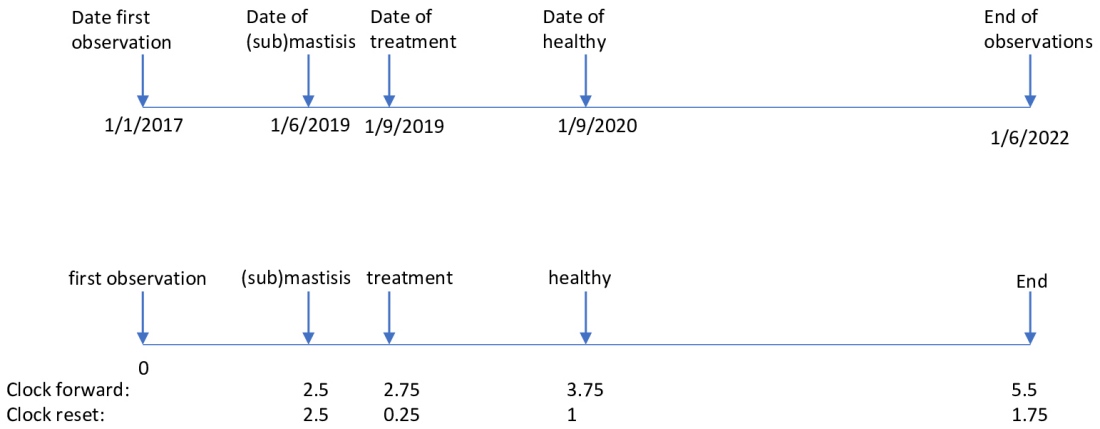



Figure 5: Illustration of the 'clock forward' and 'clock reset' approach.

In the context of Multi State Models, the "clock forward" methodology is predominantly employed. This approach consistently treats time $t$ as the duration since a patient's initial entry into a state, without being influenced by any intermediary events or transitions. This is in harmony with the foundational principles of Markov models. In contrast, the "clock reset" approach deviates from the Markov assumption, as its time perspective is intrinsically tied to the duration since the most recent state transition. The inherent simplicity of the Markov property greatly aids in calculating likelihoods within Multi State Models. By adopting the Markov assumption, transition probabilities can be effectively represented through transition intensities, as elucidated by:

$$P(s,u) = \prod_{t=s}^{u} (I + dA(t)), \tag{9}$$

where $I$ is the identity matrix and $dA(t)$ is a matrix that shows the rate of change at time $t$ (Putter et al., 2007a).

### 4.5 Cox Proportional Hazard Model with Frailty in Multi State Models

Within the literature, the Cox proportional hazard model is renowned for its ability to estimate transition probabilities. Its core principle is measuring the impact of covariates on survival

dynamics (Cox, 1972). In the framework of Multi State Models, the model is crucial in assessing the significance of various prognostic factors during state transitions.

For an individual whose characteristics are encapsulated in the covariate vector $Z$, the hazard function that defines the transition from state $g$ to $g'$ is formulated as:

$$\alpha_{gg'}(t) = \alpha_{gg',0}(t) \exp(\beta_{gg'}^{\top} Z_{gg'}). \tag{10}$$

Here, $\alpha_{gg',0}(t)$ denotes the baseline hazard linked with transitioning from state $g$ to $g'$. Simultaneously, $\beta_{gg'}$ stands as the corresponding vector of regression coefficients. In certain contexts, the model can be represented as:

$$\alpha_{gg'}(t) = \alpha_{gg',0}(t) \exp(\beta_{gg'}^{\top} Z_{gg'}). \tag{11}$$

Where $Z_{gg'}$ denotes a specific covariate vector for the transition $g \to g'$, derived from the base variable set $Z$ (P. K. Andersen et al., 1991). Using $Z_{gg',i}$ to symbolize the transition-associated covariates for an individual $i$ for the $g \to g'$ shift, the coefficients $\beta$ can be discerned by optimizing a modified Cox partial likelihood:

$$L(\beta_{gg'}) = \prod_{g \to g'} \prod_{i=1}^{n} \frac{\exp(\beta_{gg'}^{\top} Z_{gg',i})}{\sum_{j \in R_g(t_{gg',i})} \exp(\beta_{gg'}^{\top} Z_{gg',j})} \tag{12}$$

In this structure, $t_{gg',i}$ marks the event or the censoring instant for subject $i$ during the $g \to g'$ transition. Simultaneously, $R_g(t_{gg',i})$ typifies the cluster of subjects in state $g$ at instance $t$ (where $t$ represents the elapsed time post-entry into state $g$). As an illustration, if we operate within a model hosting three discrete states, a spectrum of transitions emerges. For instance, potential state transitions can flow from state 1 to either of states 1, 2, or 3. These probabilities are depicted in Equation 8. Specifically, the likelihood of an entity staying in state 1 during the time interval $(s, t]$ can be expressed as:

$$P_{11}(s, t|Z) = \exp\left( - \int_s^t (\alpha_{12}(u|Z) + \alpha_{13}(u|Z))\, du \right), \tag{13}$$

the quantity $P_{11}(s, t|Z)$ denotes the probability of remaining in state 1 over the time interval $[s, t]$ given the observation set $Z$. The exponential function, denoted by exp, encapsulates the continuous decay of probability over time due to possible transitions to other states, in this case, states 2 and 3. The integral $\int_s^t$ accumulates the effect of these transition rates from time $s$ to $t$. The terms $\alpha_{12}(u|Z)$ and $\alpha_{13}(u|Z)$ represent the transition rates from state 1 to states 2 and 3 respectively at time $u$, conditioned on the observation set $Z$. Within the framework of the

Cox model, frailty models incorporate a random effect, denoted $\omega$, to account for unobserved heterogeneity, as illustrated by the equation

$$\lambda(t|Z,\omega) = \omega\lambda_0(t)\exp(\beta' Z). \tag{14}$$

This frailty approach (FA), effectively introducing a random effects dimension to survival models, lends a refined layer to the Cox proportional hazards model. Conventional survival models may reveal correlations among survival times, particularly in grouped or clustered data, which could stem from unobserved risk factors or common attributes within clusters. These are aspects that the standard Cox model may not inherently recognize. The integration of frailty into the Cox model encapsulates these shared unobserved characteristics, offering a mechanism to adjust hazard estimations based on common latent variables. This adjustment ensures that survival estimations are more precise and aligned with the intrinsic data structure. Moreover, the inclusion of frailty facilitates a deeper comprehension of the variability in survival times among researchers. For instance, two individuals exhibiting identical observed characteristics could have varying survival probabilities due to unmeasured factors or inherent susceptibilities. The frailty term accommodates this "random effect", enabling a more personalized risk assessment. Additionally, our investigation explored the impact of varying company counts on the efficacy of the Frailty Approach (FA), probing into the notion that a higher company count may usher in diverse management practices, thereby increasing unobserved heterogeneity. As a result, a larger number of companies could potentially enhance the FA model's performance by introducing a greater volume of these unobserved variables into the data. When MSMs use CPH estimation methods, whether with or without the FA, they are known as conventional MSMs.

### 4.6 Advances in Machine Learning for Multi State Models

The advancement of machine learning has largely redefined the current landscape of Multi State Models. Tapak et al. (2018) has shed light on the integration of these paradigms, with special emphasis on the significant surge in the application of machine learning methodologies for time-to-event data analysis.

Random Survival Forests (RSF), an ensemble method adapted to survival data, is a prime example (Ishwaran et al., 2008). However, even this robust model is not without limitations. In particular, the model's limitation in handling a minimal number of observed failure events in terminal nodes is a noted impediment (Ishwaran & Kogalur, 2010b). To address such gaps,

R. Zhu and Kosorok (2012) pioneered the Recursively Imputed Survival Trees (RIST) strategy. This nonparametric imputation method iteratively refines censored observations, resulting in improved model precision. Tapak et al. (2018) further extended this concept by proposing Multi State RIST (MSRIST). This work demonstrated the superiority of MSRIST over traditional Cox hazard MSMs, but with the caveat that its application is restricted to unidirectional models.

### 4.6.1 Recursively Imputed Survival Trees for Multi State Models

Based on the work of Tapak et al. (2018), the MSRIST algorithm can be summarized in the following procedural steps:

1. Multi State Tree Model Fitting: Configuration of $M$ Extremely Randomized Multi-State Trees (ERMTs) on the preliminary dataset.
2. Conditional Transition Distribution: Computation of survival distribution for each censored entry.
3. One-Step Imputation for Censored Observations: Substitution of censored data.
4. Refit Imputed Dataset and Subsequent Analysis: Generation and fitting of imputed datasets.
5. Final Prediction: Recursive iterations to arrive at the conclusive predictions.

### 4.6.2 Random Survival Forests

Random Survival Forests (RSF), an extension of Random Forests (RF), were developed by Ishwaran et al. (2008) to adapt the RF methodology to the unique challenges of right-censored survival data. The implementation of RSF adheres to the core principles established by RF with a few key adaptations: (a) In RSF, survival trees are constructed using data sets created via bootstrapping; (b) The selection of features for splitting nodes within these trees is randomized; (c) Typically, these survival trees are allowed to grow to significant depths; and (d) The final ensemble of the survival forest is derived by averaging the terminal node statistics (TNS) across all trees. This approach allows RSF to effectively handle and analyze survival data.

The `randomForestSRC` R package offers this technique through the `rfsrc()` function. We can adjust this function using various parameters that influence tree building (Ishwaran & Kogalur, 2021). Choosing these parameters can significantly influence prediction results (Ishwaran

& Kogalur, 2021). The software documentation emphasizes the `nodesize` and `ntree` parameters, which represent the unique cases in each final node and the number of trees utilized. Fine-tuning these parameters can enhance the model's predictive ability.

## 4.7 Prediction performance

The purpose of this thesis is to compare the Cox Proportional Hazards model and machine learning techniques in Multi State Models, with the primary goal of evaluating their ability to accurately predict survival during state transitions in cows. The assessment of predictive accuracy is performed using measures of calibration. Calibration assesses the consistency between predicted and observed survival probabilities. If a model is well calibrated, its predicted survival probabilities will closely match the observed probabilities in all subgroups. Discrimination, on the other hand, assesses a model's ability to distinguish between subjects based on their survival probabilities. In survival analysis, the concordance index ($C$-index) is often used to measure this. Lammens (2014) shows that the $C$-index and the Brier Score (BS) are the most robust and efficient way to compare the predictive power between the MSRIST and CH models. In addition, the prediction accuracy can be evaluated using the prediction error, where the Brier Score is an important method for this purpose. It assesses the prediction error with respect to survival beyond a predetermined time point. To provide a broader perspective, the Integrated Brier Score (IBS) functions as a thorough measure of prediction error. This section focuses on the calculation of the $C$-index and the IBS and examines their use in both the CPH model - with and without a FA - and random survival forest models.

### 4.7.1 Concordance Index

The $C$-index serves as a measure to assess the concordance between predicted and observed survival times. As described in Harrell et al. (1996), this index analyzes pairs of cows from the data to determine if the cow with the highest predicted survival time actually lived longer. Pairs are excluded if both cows are censored or if the cow with the shorter observed time is censored, because in such cases, it is unclear who lived longer. Additionally, pairs in which both cows die simultaneously are excluded. To calculate the $C$-index we let $T_i$ and $T_j$ be the observed survival times for cows $i$ and $j$ respectively:

- Count 1 for every pair $(i, j)$ where $T_i \neq T_j$ if the cow with the longer predicted survival time indeed survived longer.

- Count 0.5 for pairs $(i, j)$ where $T_i \neq T_j$ when both cows have identical predicted survival times, or for $T_i = T_j$ if the censored cow has a longer predicted survival time.

The $C$-index is then defined as the ratio of concordant pairs

$$C\text{-index} = \frac{\text{Number of concordant pairs}}{\text{Total number of informative pairs}}. \tag{15}$$

The optimal value for the $C$-index stands at 1, indicative of perfect prediction. A score of 0.5 implies a lack of predictive ability, same as random guessing. Within certain models, such as the CPH model, a direct relationship exists between estimated survival times and survival probabilities. This allows the use of prognostic indexes in evaluating concordance.

When using the Random Survival Forest (RSF) algorithm within MSMs, the $C$-index is estimated through ensemble transition probabilities. The term "ensemble" refers to an average over the whole forest, capturing the nuances involved in state transitions. For a given transition from state $g$ to state $g'$, based on the covariate vector $x_i$, the ensemble transition probability is given by the equation:

$$\hat{P}^*_{g,g',e,i} = \sum_{j=1}^{n} H^*_e(T_j | x_i, g \to g'),$$

where $T_j$ represents the times in which transition probabilities are summed for $j = 1, ..., n$, $n$ denotes the total number of times being considered, and $e$ represents a specific event or state transition such as moving from healthy to sick or from sick to death. The function $H^*_e(T_j | x_i, s \to s')$ denotes the ensemble average cumulative hazard function (CHF) for the specific transition, obtained by averaging across all trees in the forest. For a more unbiased estimate, Ishwaran and Kogalur (2010b) utilize the out-of-bag (OOB) ensemble transition probability, represented by the equation:

$$\hat{P}^{**}_{g,g',e,i} = \sum_{j=1}^{n} H^{**}_e(T_j | x_i, g \to g').$$

Pre-selected times for evaluating transitions are represented by $t_{o1}, ..., t_{om}$. If the sum of $H^{**}_e(t_{ol} | x_i, g \to g')$ for cow $i$ is greater than the sum of $H^{**}_e(t_{ol} | x_j, g \to g')$ for cow $j$, then there is a higher probability of transition for cow $i$ compared to cow $j$.

The $C$-index based on the OOB ensemble transition probabilities is termed as $C^{**}$, and its corresponding OOB prediction error is

$$\text{err}^{**} = 1 - C^{**}. \tag{16}$$

This prediction error gauges the effect of a specific variable randomization on assessing variable importance in the RSF for multi-state models.

### 4.7.2 Integrated Brier Score

The flexibility of the IBS is especially useful in studying survival analysis and MSMs. It is essential in determining the accuracy of transition probabilities between different states during a specified time period. Monitoring the progress of individuals moving from healthy to sick or recovered states requires exceptional precision. The main aim is to predict these transitions with the highest degree of accuracy. In the realm of MSMs, the IBS formula is an elevated version of the standard BS. The difference is that instead of measuring simple outcomes, it compares the expected and actual transitions between states. To use the IBS formula for an MSM with $G$ states and $U$ timepoints, where cows move forward from a start timepoint of $u = 0$, the following formula can be applied:

$$\text{Integrated Brier Score} = \frac{1}{U}\sum_{u=1}^{U}\left(\frac{1}{KN(U-1)}\sum_{g=1}^{G}\sum_{i=1}^{N}(P_{gig'} - O_{kig'})^2\right). \tag{17}$$

In this equation, the symbol $K$ refers to the total number of states, $N$ represents the total number of participants, $U$ stands for the total number of time points, $P_{gig'}$ is the estimated likelihood of a participant moving from state $g$ to $g'$ at time point $u$, and $O_{gig'}$ shows the actual transition measurement, where 1 denotes that the transition occurred and 0 denotes the opposite, for a participant in state $g$ at time point $u$. The IBS is calculated by adding up the differences between the expected and real chances for movement in every state, person, and time period. This total is then divided by all possible movements and time periods. Just like the original, a lower IBS in MSMs means increased accuracy.

### 4.7.3 Likelihood for the Conventional Multi State Model

Likelihood is a fundamental concept for assessing the performance of statistical models, including Conventional Multi State Models (MSMs). MSMs extend traditional survival analysis by allowing for multiple states beyond the binary outcome. For instance, in a livestock context, an animal can transition through states such as “healthy”, “sick”, and “dead”. The likelihood for an MSM is based on the transition intensities, or probabilities, for each possible state transition. Using $q_{gg'}(t|\theta)$ to represent the transition intensity from state $g$ to state $g'$ at time $t$, and $P_{gg'}(t|\theta)$ as the probability of being in state $g'$ at time $t$, given starting in state $g$ at time $t = 0$, the likelihood function is given by

$$L(\theta) = \prod_{n=1}^{N} \prod_{g} \prod_{g'} \left(q_{gg'}(t|\theta)\right)^{d_{gg'}^{(n)}(t)} \left(P_{gg'}(t|\theta)\right)^{1-d_{gg'}^{(n)}(t)} . \tag{18}$$

In this equation, $\theta$ denotes the model parameters, and $d_{gg'}(t)$ is an indicator function that equals 1 if a transition from $g$ to $g'$ is observed at time $t$, and 0 otherwise. Comparing likelihoods across different models can be challenging due to varying model structures and assumptions. For comparative evaluations, alternative metrics like the $C$-index or Integrated Brier Score (IBS) may be more effective. Considering computational difficulties and precision concerns with multiplying many likelihood values, especially in large datasets, the negative log-likelihood is often used:

$$-\log L(\theta) = -\sum_{n=1}^{N} \sum_{g} \sum_{g'} \log \left( q_{gg'}(t|\theta)^{d_{gg'}(t)} P_{gg'}(t|\theta)^{1-d_{gg'}(t)} \right) . \tag{19}$$

This approach sums the negative log-likelihoods over all individuals (N) and all observed transitions between states $g$ and $g'$, which is numerically more stable and provides a clearer assessment of the model's fit.

In equations above, the double products $\prod_g \prod_{g'}$ and the double sums $\sum_g \sum_{g'}$ iterate over all possible start and stop states.

To ensure computational stability and precision, especially when dealing with large datasets, we often utilize the negative log of the normalized likelihood. The normalization of the likelihood function by the number of observations adjusts for different sample sizes, allowing for meaningful comparisons between datasets. The normalized negative log-likelihood is given by the following equation:

$$-\log L_{\text{norm}}(\theta) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{g} \sum_{g'} \log \left( q_{gg'}(t_n|\theta)^{d_{gg'}(t_n)} P_{gg'}(t_n|\theta)^{1-d_{gg'}(t_n)} \right) . \tag{20}$$

Here, $N$ stands for the total number of observations, ensuring that the likelihood is averaged over all observations. This equation compiles the negative log-likelihoods across all individuals and all observed transitions between states $g$ and $g'$. By normalizing the sum by $N$, we obtain a measure of fit per observation, which is particularly useful when comparing models across datasets of varying sizes. This approach not only provides a numerically stable solution but also a consistent metric for model comparison.

# 5 Results

In this section, we present the results and interpretations for the CPH, CPH-FA, and MSRIST models. We dive into a comparison of the unobserved heterogeneity between these models by varying the number of farms and observing variations in predictions based on the $C$-index, IBS, and likelihood. Specifically for the MSRIST model, we performed hyperparameter tuning and will detail the parameters used. The framework for this model comparison can be found in Section 5.1. In Section 5.2, we show the difference in results between different farm sizes. Finally, Section 5.3 examines the effects of Antibiotica and Quick-Extra treatments on a cow's state transitions.

## 5.1 Methodology and Hyperparameter Optimization for Model Comparison

In order to fully comprehend the dynamics of this study, this section outlines the specific settings and computations utilized. The main goal was to conduct a comprehensive comparison between traditional MSMs and the newer MSRIST. A crucial aspect of this comparison focused on determining how these models perform under different levels of farm quantities. Based on literature sources, we hypothesized that as the number of farms increases, the likelihood of encountering unexpected heterogeneity also increases due to different management practices. In accordance with our hypothesis, we asserted that the CPH-FA and MSRIST would demonstrate better output in situations with greater variability, as opposed to those without FA. To test this, we established a range of farm quantities, starting at a minimum of 25 farms and gradually progressing to a maximum of 1000 farms. Our analysis involved several intermittent steps, specifically examining farms at $n = 50$, $n = 100$, $n = 250$, $n = 500$ and $n = 1000$.

Our MSRIST research began with an essential step: optimizing hyperparameters. Recognizing the critical role that hyperparameters play in the performance of machine learning models, we embarked on the meticulous process of tuning these parameters. The tool we chose was Random Survival Forest, a powerful method for achieving improved performance. We conducted a thorough tuning process that took a total of 36 hours. We have thoroughly documented the results of this process, the optimal parameters, in Table 7. For readers interested in a more detailed analysis, more precise graphs are provided in Appendix C Figure 10.

After tuning our models, the focus shifted to objectively evaluating their performance when implemented in different numbers of companies. We also evaluated their performance at each transition.

Table 7: Hyperparameters after tuning

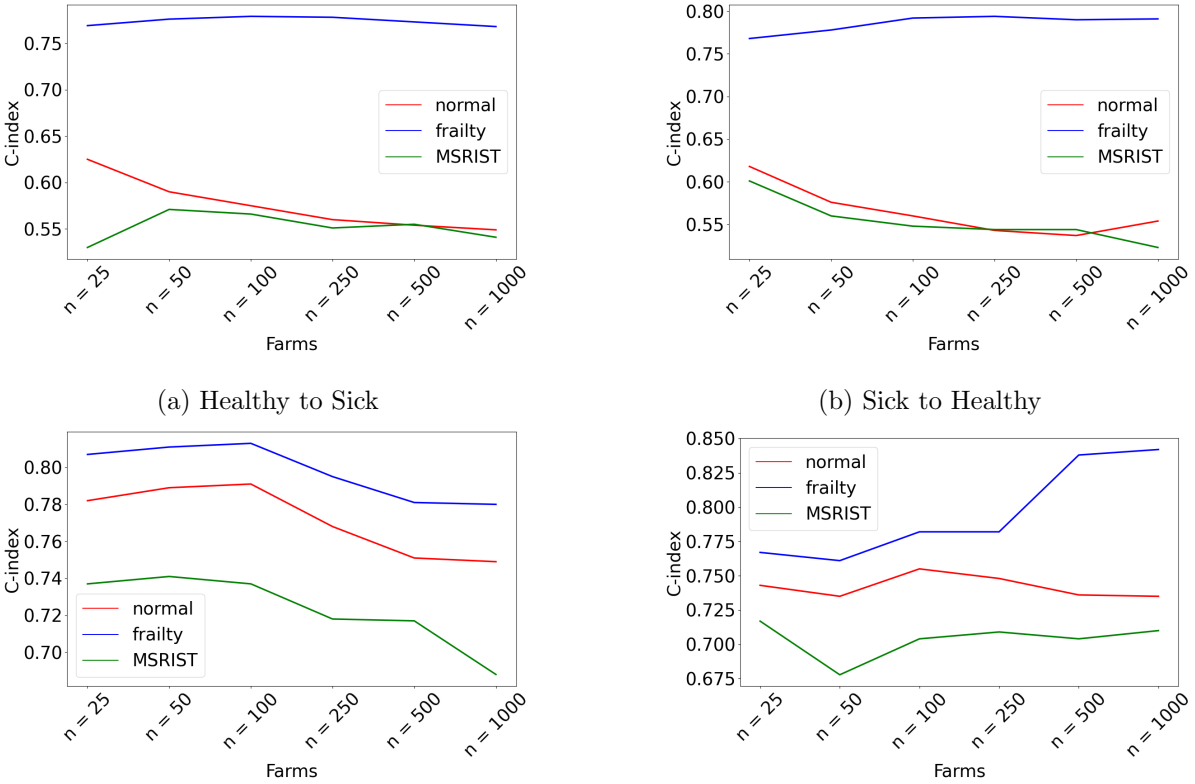
| Transition | No. tree | Node-size | OOB Error rate |
|---|---|---|---|
| Healthy to Sick | 200 | 6 | 0.459 |
| Sick to Healthy | 100 | 4 | 0.474 |
| Healthy to Died | 100 | 3 | 0.314 |
| Sick to Died | 100 | 6 | 0.298 |

### 5.2 Comparative Model Performance

In our research, we aim to examine the impact of an increased number of farms on the performance of the Cox Proportional Hazard (CPH) model, the Cox Proportional Hazard - Frailty Approach (CPH-FA), and the MSRIST model. This analysis will consider farms of different sizes to determine if a larger number of farms, which may have more unobserved heterogeneity, affects model performance. We will focus on evaluating two primary performance indicators: the $C$-index, which assesses each model's ability to accurately discriminate and rank individual risks, and the Integrated Brier Score (IBS), which measures predictive accuracy. The analysis examines whether the performance of the models differs as the number of farms increases, possibly due to better handling of unobserved heterogeneity. A numerical overview of the results is presented in the Appendix B Table 11.

By examining Figure 6, we can clearly see that the CPH-FA model consistently outperforms its regular CPH counterpart across all transitions. This highlights the critical importance of identifying and accounting for hidden variability within the data set. These often overlooked or unobservable variations can have a significant impact on model predictions. The frailty approach effectively captures these nuances, thereby improving the accuracy of the model. Moreover, as the dataset expands to include information from more companies, a pattern emerges. The performance gap between the standard CPH model and its frailty-enhanced version widens. This suggests that the CPH-FA model is more adaptable and remains more relevant, especially as the dataset grows to include a larger number of farms. To validate the observed differences, we conducted both the Likelihood Ratio Test (LRT) and the Wald test. The results of both tests confirm our observations and indicate that the differences between the models are statistically significant, with p-values $<0.001$. This statistical evidence supports the superior performance of the CPH-FA model. In addition, our analysis shows that for small values of $n$, most transitions tend to produce better results. However, it's worth noting that the model's performance for the

"healthy to sick" and "sick to healthy" transitions is somewhat worse than for other transitions. A possible reason for this discrepancy could be the inherent complexity and multifaceted nature of health transitions. The transition from a "healthy to sick" state, or vice versa, may be influenced by a variety of factors that may be difficult for the model to capture comprehensively such as genetic variations. In addition, our observations suggest that the MSRIST algorithm tends to perform slightly worse based on the $C$-index than other algorithms. This relative underperformance may be due to the underlying assumptions of the algorithm or its sensitivity to certain data distributions.




(a) Healthy to Sick

(b) Sick to Healthy

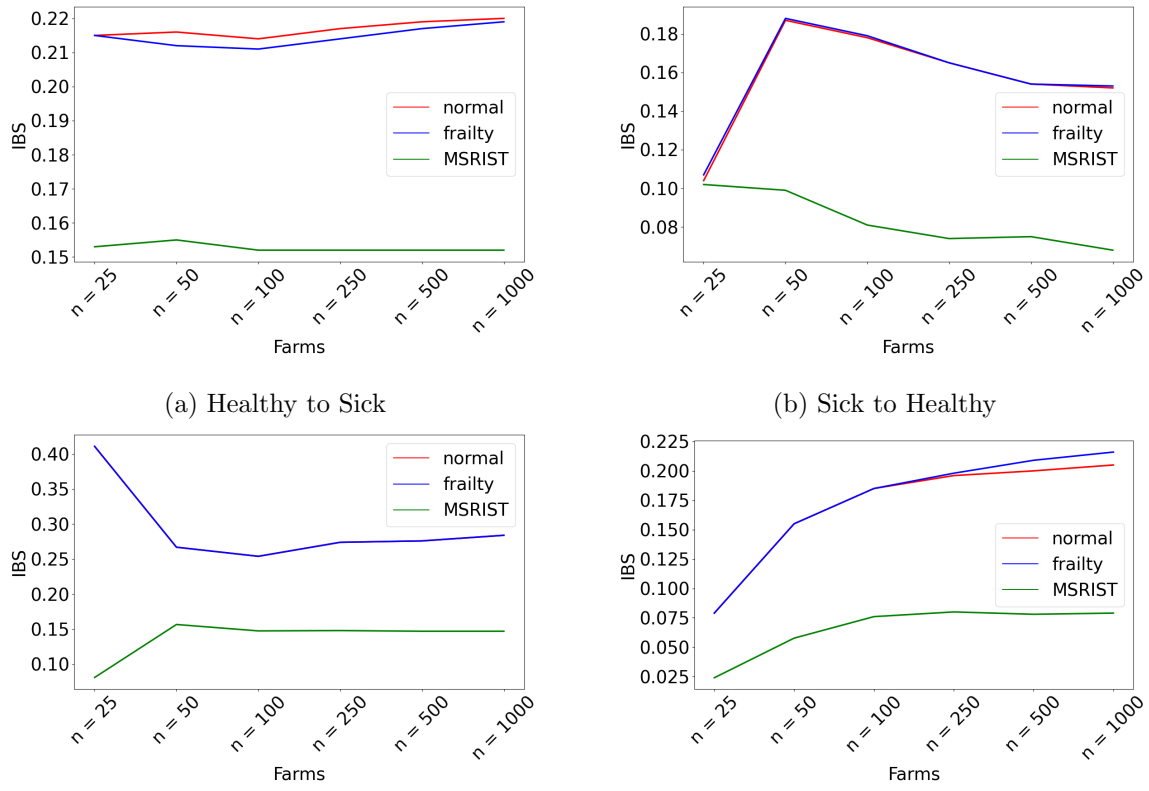
(c) Healthy to Death

(d) Sick to Death

Figure 6: C-Index values across different transitions and farms counts

In evaluating the performance metrics of survival prediction models, IBS serves as a critical determinant, with lower values indicating superior predictive ability. Figure 7 provides a visual representation of our findings across different transitions and farms counts. When the number of farms increases, there is a noticeable rise in the IBS value for the transition from "Sick to Death". This trend suggests a decline in prediction accuracy, potentially due to the added complexity and diversity in larger datasets, which may challenge the model's ability to consistently predict this particular transition. The CPH (Cox Proportional Hazards) model and its Frailty Approach

variant, CPH-FA, show very close performance metrics. The similarity in their outcomes might suggest that the frailty component, which is intended to account for unobserved heterogeneity, does not significantly impact the hazard estimation in this context. On the other hand, the MSRIST (Multistate Recursively Imputed Survival Tree) method consistently surpasses both the CPH and CPH-FA models in predictive power. This could be because MSRIST employs a more sophisticated approach to handle the multistate nature of survival data, effectively dealing with censoring and the recursive partitioning of the data. Its superior performance is reflected in the lower IBS scores, indicating a more precise estimation of survival probabilities. Although direct comparisons should be approached with caution due to potential variations in the nature of the transitions and data sets between studies, it is observed that our MSRIST algorithm has a lower IBS then the CPH variant, which is similar to the results of Tapak et al. (2018). In addition, the $C$-index of the CPH-FA in our study is in the same range as that reported by Tapak et al. (2018), suggesting similarities in predictive ability. These findings provide a preliminary understanding of the ability of our models, particularly the MSRIST algorithm, to estimate survival probabilities.




(a) Healthy to Sick

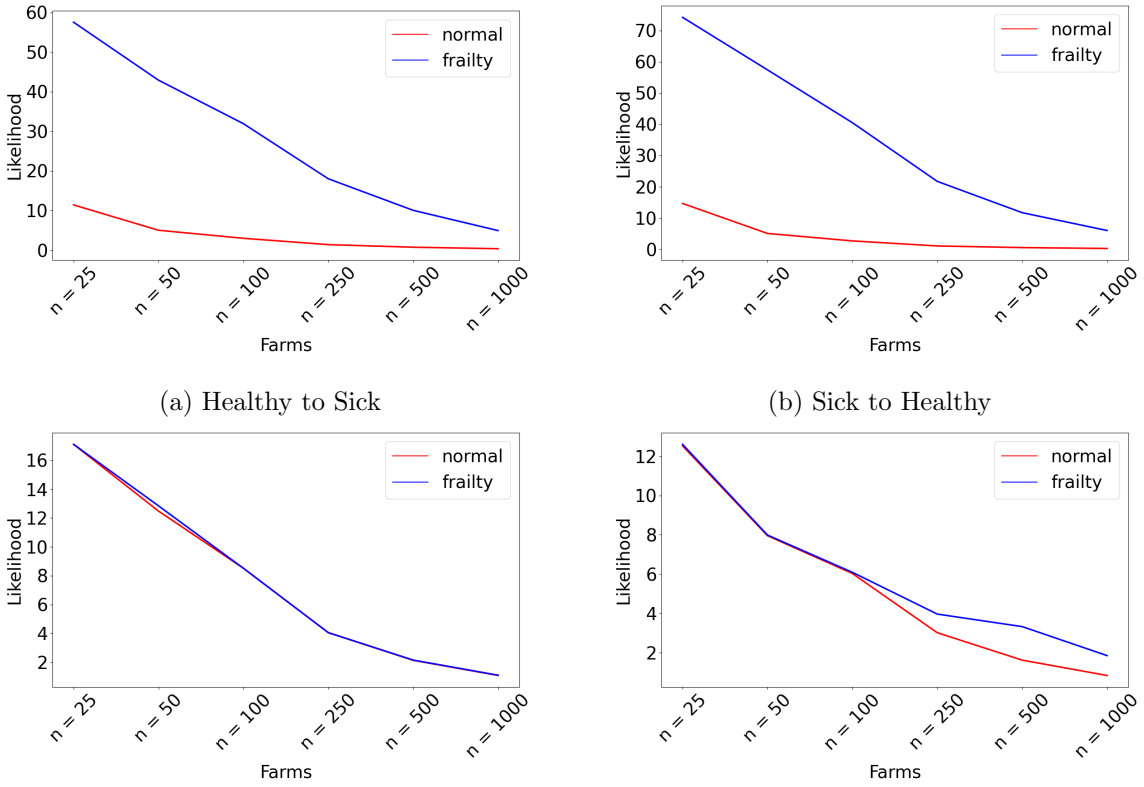
(b) Sick to Healthy

(c) Healthy to Death

(d) Sick to Death

Figure 7: IBS values across different transitions and farms counts

After understanding the model's performance through the lens of the $C$-index and IBS, our attention now turns to the likelihood values. Likelihood offers a view of the model's ability to fit the data, making it a critical measure in this comparative analysis. A visualization of the likelihood values is available in Figure 8, revealing the comparative performance of the two models.




(a) Healthy to Sick

(b) Sick to Healthy

(c) Healthy to Death

(d) Sick to Death

Figure 8: Normalized Likelihood values across different transitions and farms counts

After examining the data, we found that the likelihood values align with the concordance index. The frailty model consistently produces higher likelihood values, indicating a better fit to the data compared to the standard CPH model. This trend becomes more pronounced as the dataset expands in terms of company counts. One notable example of this is in the "Healthy to Sick" transition, where the frailty model's ability to account for unobserved heterogeneity becomes increasingly crucial.

### 5.3 Treatment Analysis: Effects of Antibiotics and Quick-Extra on State Transitions

When analyzing the effects of treatments on cow transitions, it is important to understand how different factors play a role in these transitions. In particular, the two treatments of interest in this analysis are antibiotic use and QE. Given the widespread use and impact of these treatments, understanding their effects can provide valuable insights into their efficacy and potential side effects.

In addition, age is another important factor included in this analysis. Age is a natural determinant of health and vitality. As cows age, their immune response and overall physiological state may change, making them potentially more susceptible to certain diseases or responding differently to treatments than younger cows. Incorporating age into our analysis provides a more holistic view, ensuring that any observed effects are not simply due to the natural aging process, but are actually influenced by the treatments being studied.

The results derived from the different transition models, specifically using the CPH and CPH-FA methods, are presented in the Table 8. Within Table 8, the key coefficient is labeled as Exp(Coef). This represents the exponential coefficient and can be interpreted as the multiplicative change in hazard for a one-unit increase in the predictor variable. For example, a Exp(Coef) value greater than 1 indicates an increased risk, while a value less than 1 indicates a protective effect or decreased risk.

For the "healthy to sick" transition, both QE and antibiotics are associated with increased risk, with the effect of QE being more significant. This makes sense when considering the mechanism of action of QE, which may initially lead to higher SCC due to biofilm disruption and subsequent bacterial release before overall health improves. Consequently, there is a logical basis for the observation that QE is associated with a higher risk of health deterioration in the short term. Antibiotics, on the other hand, typically have a direct and immediate effect on infection without an initial increase in SCC. At the same time, age shows a positive correlation, suggesting that the risk of this transition increases as cows get older. When looking at the transition "Sick to Healthy", a similar trend emerges, albeit with stronger effects. Interestingly, the transition from healthy to dead shows a protective effect attributed to QE, in contrast to antibiotics, which shows increased risk. This trend persists in the transition from sickness to death, underscoring the increased risk associated with both treatments, with particular emphasis on the influence of antibiotics. These observations provide invaluable insights into the efficacy

Table 8: Comparative Analysis of Health Transition Probabilities Using CPH and CPH-FA Models

| | | CPH | | CPH-FA | |
|---|---|---|---|---|---|
| **Transition** | **Variable** | **Exp(Coef)** | **Std Error** | **Exp(Coef)** | **Std Error** |
| **Healthy to Sick** | QE | 1.531 | 0.079 | 1.488 | 0.085 |
| | Antibiotica | 1.159 | 0.075 | 1.123 | 0.080 |
| | Age | 1.067 | 0.003 | 1.070 | 0.004 |
| **Sick to Healthy** | QE | 2.199 | 0.064 | 2.297 | 0.071 |
| | Antibiotica | 1.729 | 0.058 | 1.782 | 0.066 |
| | Age | 1.041 | 0.004 | 1.040 | 0.004 |
| **Healthy to Death** | QE | 0.520 | 0.164 | 0.520 | 0.164 |
| | Antibiotica | 3.924 | 0.166 | 3.927 | 0.166 |
| | Age | 1.153 | 0.014 | 1.153 | 0.014 |
| **Sick to Death** | QE | 1.170 | 0.145 | 1.224 | 0.147 |
| | Antibiotica | 4.976 | 0.141 | 5.513 | 0.143 |
| | Age | 1.300 | 0.014 | 1.316 | 0.014 |

and potential hazards associated with QE and antibiotics. This is particularly important when considering the age of the cows.

When analyzing the MSRIST algorithm regarding state transitions, it becomes clear that different variables have different degrees of influence. In contrast to the CPH and CPH-FA models that enable hazard functions' derivation, the MSRIST algorithm emphasizes the significance of certain variables for predicting transitions. The regularity of these variables' effect is congruent with results from CPH and CPH-FA models. Table 9 shows which variables are most important according to the MSRIST algorithm.

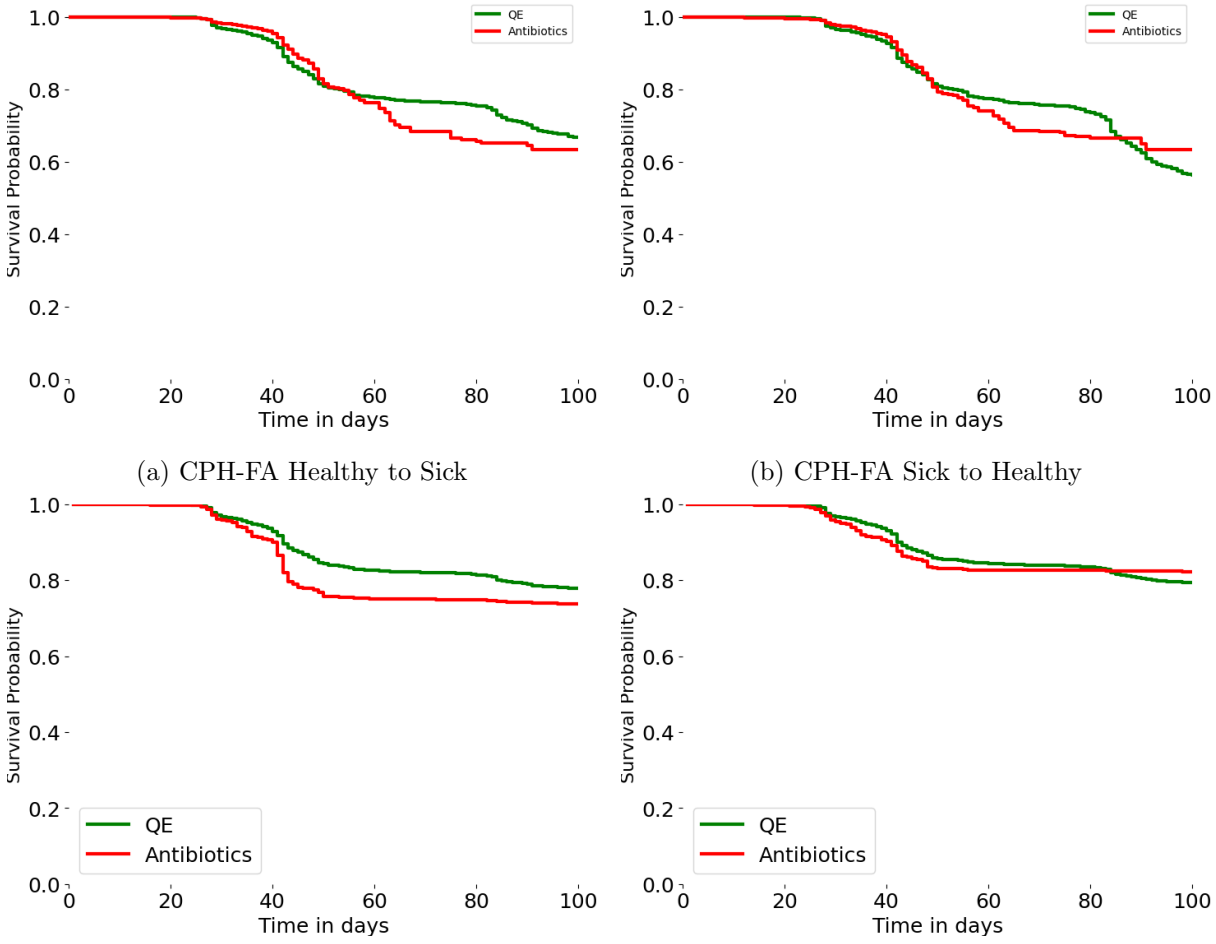
As evidenced in the table, "age" consistently exhibits a clear and logical impact on all transitions. While "antibiotics" holds significant weight, particularly in mortality-related transitions "Healthy to Dead" and "Sick to Dead", "QE" is also an influencing factor but with generally lower feature importance across most transitions.

The most common way to visualize time-to-event data is a survival curve. The curves start at 100% and go down as events accumulate. Specifically, the curves show the probability of "surviving" - that is, avoiding the event of interest - over a given period of time. A close analysis

Table 9: Feature Importance of MSRIST model

| Transition | Variable | Feature Importance |
|---|---|---|
| **Healthy to Sick** | QE | 0.158 |
| | antibiotics | 0.137 |
| | age | 0.275 |
| **Sick to Healthy** | QE | 0.165 |
| | antibiotics | 0.152 |
| | age | 0.283 |
| **Healthy to Dead** | QE | 0.072 |
| | antibiotics | 0.298 |
| | age | 0.341 |
| **Sick to Dead** | QE | 0.111 |
| | antibiotics | 0.316 |
| | age | 0.358 |

of the survival plots for both the CPH-FA approach and the MSRIST algorithm reveals notable similarities and nuanced differences, as shown in Figure 9. The survival curves (indicating the probability of not experiencing a state transition) for both methods show comparable trajectories, particularly in how they decrease over time. However, upon closer inspection, the MSRIST algorithm shows a slightly faster decline in its survival curves than the CPH-FA approach. This steeper slope implies that state transitions may occur at a higher rate or earlier in time for MSRIST, suggesting that MSRIST may be more suited to certain subtleties or peculiarities in the data. The CPH-FA approach, on the other hand, may be more conservative in detecting these transitions. Despite these differences, it's important to note that the overall shapes and trends seen in the results of both algorithms show similarities between the two algorithms, or within the same dataset. This similarity is evident in Figure 9, which likely shows either overlapping or closely parallel survival curves for the two methods. Such visual evidence supports the notion that even if they diverge in terms of sensitivity or detection rate, their underlying behaviors and insights remain largely aligned.




(a) CPH-FA Healthy to Sick

(b) CPH-FA Sick to Healthy

(c) MSRIST Healthy to Sick

(d) MSRIST Sick to Healthy

Figure 9: Survival plots for different algorithms, starting at a 100% survival probability indicating all subjects are initially alive, with subsequent declines reflecting event occurrences over time.

# 6 Conclusion and Discussion

In the research conducted, our primary objective was to address the following research question: "How do Quick-Extra treatment outcomes for mastitis compare to antibiotics when evaluated using MSMs and MSRIST, and is it possible for machine learning to enhance the precision of conventional MSMs?". To this end, we performed comprehensive comparisons between the CPH, CPH-FA, and MSRIST methods based on criteria such as IBS, and $C$-index. By analyzing the coefficients derived from our estimated models, we sought to uncover the distinct effects of different treatments on state transitions. Our results suggest that the MSRIST methodology outperforms the CPH-FA and CPH when assessing the IBS. Furthermore, the CPH-FA and MSRIST models show congruent directions with respect to survival probability estimates. Interestingly, we observed that the effects of antibiotics and Quick-Extra (QE) on state transitions are largely similar.

The CPH-FA method highlighted the importance of capturing unobserved heterogeneity, especially when accounting for variability in farm management practices. As the number of farms increased, the performance of the CPH-FA model, as measured by the $C$-index and likelihood, consistently exceeded that of the standard CPH model. This trend was evident as the number of farms increased from 25 to 1000; the $C$-index improved from 0.625 (for CPH) to 0.769 (for CPH-FA) at 25 farms, and from 0.549 (for CPH) to 0.768 (for CPH-FA) at 1000 farms. This suggests that the CPH-FA model is better able to account for the additional unobserved heterogeneity introduced by the inclusion of more farms. While both methods showed comparable results on the IBS, indicating similar predictive accuracy over time, the improved $C$-index for the CPH-FA model implies a superior ability to rank survival times in the presence of unobserved factors under different farm management conditions. The MSRIST technique showed better results than the CPH-FA method in terms of the IBS. IBS is a comprehensive indicator of prediction accuracy over time in survival analysis. This advantage can be linked to MSRIST's ability to handle complex data patterns and use decision trees to identify intricate interactions that may not be straightforward. MSRIST's ability to fill in missing data reduces the impact of incomplete datasets, leading to more reliable analysis. Although MSRIST performs well in this regard, it does not outperform CPH-FA in terms of the $C$-index, a commonly used measure of the accuracy of risk scores. This implies that although MSRIST can handle the intricacies of survival data and missing facts, the CPH-FA model's strategy for taking into account undetected diversity through frailties might offer a advantage

in the context of the $C$-index. The $C$-index assesses how well the model can rank individuals by risk, a task where the correction for unmeasured factors via frailties can be particularly advantageous. Choosing the right model based on the data's specific features and analytic objectives is vital.

From our research of the effects of antibiotic versus Quick-Extra (QE) treatments on cattle, we have uncovered some interesting patterns. The comparative analysis suggests that, at first glance, both treatments influence state transitions similarly. However, a closer look reveals differences. One striking observation is the increased likelihood of antibiotic-treated cows entering the "death" state. This trend can be attributed to a number of reasons. First, antibiotics are often reserved for more severe cases, mainly because of the impact on milk production. In addition, economic considerations force farmers to send antibiotic-treated cows to the slaughterhouse (Manyi-Loh et al., 2018). In contrast, while QE stands out as a natural remedy that can be used earlier in the treatment timeline without compromising milk quality or posing a risk to consumers, our findings suggest a more nuanced picture of its effects. In particular, there appears to be a higher risk of transition from a healthy to a sick state in cows treated with QE. This observation may be attributed to the mechanism by which QE disrupts the biofilm, potentially releasing a higher load of bacteria into the system and consequently leading to an increase in SCC. This is in contrast to antibiotics, which tend to kill bacteria directly and may result in an immediate reduction in SCC. However, in the long term, our data suggest that cows treated with QE may have a greater likelihood of recovering from disease and moving from a diseased to a healthy state. This could mean that despite the initial increase in SCC, QE may ultimately have a beneficial effect on cow health over a longer period of time. However, a structured experiment would be ideal to obtain a more accurate comparative assessment between the two treatments. This would involve a controlled design with distinct control and treated groups of cows, all housed in the same farm environment. Within this design, some cows would receive QE while others would receive antibiotics, allowing for a direct comparison. Another important aspect that emerges from our analysis, which is particularly evident in the CPH-FA approach, is the strong role of unobserved heterogeneity. This factor has a significant impact and can substantially bias the results of state transitions. It's also worth noting that due to the different timing of treatment application between antibiotics and QE, a direct one-to-one comparison of their effects becomes challenging. Based on the datasets we acquired from CRV and the Dutch government, we identify similar patterns. However, more research is needed in this area to draw more definitive conclusions.

# References


Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. Springer.

Aalen, O. O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, *5*, 141–150.

Abebe, R., Hatiya, H., Abera, M., Megersa, B., & Asmare, K. (2016). Bovine mastitis: Prevalence, risk factors and isolation of staphylococcus aureus in dairy herds at hawassa milk shed, south ethiopia. *BMC Veterinary Research*, *12*, 270. https://doi.org/10.1186/s12917-016-0905-3

Abranches, J., Miller, J. H., Martinez, A. R., Simpson-Haidaris, P. J., Burne, R. A., & Lemos, J. A. (2011). The collagen-binding protein cnm is required for streptococcus mutans adherence to and intracellular invasion of human coronary artery endothelial cells. *Infection and immunity*, *79*(6), 2277–2284. https://doi.org/10.1128/IAI.00767-10

Andersen, K., & Keiding, N. (2002). Multi-state models for event history analysis. *Stat Methods Med Res*, *11*, 91–101.

Andersen, P. K., Hansen, L. S., & Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process. *Scandinavian Journal of Statistics*, *18*, 153–166.

Andersen, P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, *11*(2), 91–115.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

Bobbio, A., Puliafito, A., Telek, M., & Trivedi, K. S. (1998). Recent developments in non-markovian stochastic petri nets. *Journal of Circuits, Systems and Computers*, *8*(1), 119–158. https://doi.org/10.1142/S0218126698000067

Cheng, W. N., & Han, S. G. (2020). Bovine mastitis: Risk factors, therapeutic strategies, and alternative treatments — a review. *Asian-Australasian Journal of Animal Sciences*, *33*(11), 1699–1713. https://doi.org/10.5713/ajas.20.0156

Cox, D. R. (1972). Regression models and life-tables [Citations: 10,181]. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–220. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

CRV. (2023). Crv official website.

Davies, D. G., Parsek, M. R., Pearson, J. P., Iglewski, B. H., Costerton, J. W., & Greenberg, E. P. (2017). The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science*, *280*(5361), 295–298.

Down, P., Bradley, A. J., Breen, J., Hudson, C., & Green, M. J. (2019). Current management practices and interventions prioritised as part of a nationwide mastitis control plan. *Journal of Dairy Research*, *86*(1), 1–8. https://doi.org/10.1017/S0022029918000884

Duchateau, L., & Janssen, P. (2008). *The frailty model*. Springer Science  Business Media.

Gomes, F., & Henriques, M. (2016). Control of bovine mastitis: Old and recent therapeutic approaches. *Current Microbiology*, *72*, 377–382. https://doi.org/10.1007/s00284-015-0958-8

Gupta, P., Sarkar, S., Das, B., Bhattacharjee, S., & Tribedi, P. (2019). Biofilm, pathogenesis and prevention–a journey to break the wall: A review. *Archives of Microbiology*, *201*(1), 1–15.

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387.

Hogeveen, H., Steeneveld, W., & Wolf, C. A. (2019). Production diseases reduce the efficiency of dairy production: A review of the results, methods, and approaches regarding the economics of mastitis. *Annual Review of Resource Economics*, *11*, 289–312. https://doi.org/10.1146/annurev-resource-100518-093954

Hossain, M. A., Paul, S., Hossain, M. R., Islam, M. R., & Alam, M. S. (2017). Bovine mastitis and its therapeutic strategy doing antibiotic sensitivity test. *Austin Journal of Veterinary Science and Animal Husbandry*, *4*, 1030.

Hothorn, T., Lausen, B., Benner, A., & Radespiel-Tröger, M. (2004). Bagging survival trees. *Statistics in Medicine*, *23*(1), 77–91. https://doi.org/10.1002/sim.1628

Hougaard, P. (1999). Multi-state models: A review. *Lifetime Data Analysis*, *5*(1), 83–119.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer.

Isensee, C., Wickborn, F., & Horton, G. (2023). Training hidden non-markov models [E-mail: {claudia,fabian,graham}@sim-md.de]. *Institut für Simulation und Graphik, Otto-von-Guericke-Universität Magdeburg*.

Ishwaran, H., & Kogalur, U. (2021). Fast unified random forests for survival, regression, and classification (rf-src), r package version 2.11.0.

Ishwaran, H., & Kogalur, U. B. (2010a). Consistency of random survival forests. *Statistics Probability Letters*, *80*(13), 1056–1064. https://doi.org/10.1016/j.spl.2010.02.017

Ishwaran, H., & Kogalur, U. B. (2010b). Consistency of random survival forests. *Statistics & Probability Letters*, *80*(13), 1056–1064. https://doi.org/10.1016/j.spl.2010.02.011

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3), 841–860. https://doi.org/10.1214/08-AOAS169

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for r. *Journal of Statistical Software*, *38*(8), 1–28.

Khan, M., & Khan, A. (2006). Basic facts of mastitis in dairy animals: A review. *Pak Vet J*, *26*, 204–208.

Kibebew, K. (2017). Bovine mastitis: A review of causes and epidemiological point of view. *Journal of Biology, Agriculture and Healthcare*, *7*, 1–14.

Król, J., Brodziak, A., Litwińczuk, Z., & Litwińczuk, A. (2013). Effect of age and stage of lactation on whey protein content in milk of cows of different breeds. *Polish Journal of Veterinary Sciences*, *16*, 395–397. https://doi.org/10.2478/pjvs-2013-0055

Lakew, B. T., Fayera, T., & Ali, Y. M. (2019). Risk factors for bovine mastitis with the isolation and identification of streptococcus agalactiae from farms in and around haramaya district, eastern ethiopia. *Tropical Animal Health and Production*, *51*, 1507–1513. https://doi.org/10.1007/s11250-019-01838-w

Lammens, V. (2014). *Estimating the prediction error in multistate models*.

Le-Rademacher, J. G., Therneau, T. M., & Ou, F.-S. (2022a). The utility of multistate models: A flexible framework for time-to-event data. *Current Epidemiology Reports*, *9*(3), 183–189. https://doi.org/10.1007/s40471-022-00291-y

Le-Rademacher, J. G., Therneau, T. M., & Ou, F.-S. (2022b). The utility of multistate models: A flexible framework for time-to-event data [Open Access]. *Current Epidemiology Reports*, *9*, 183–189. https://doi.org/10.1007/s40471-022-00347-2

Limnios, N., & Oprisan, G. (2001). *Semi-markov processes and reliability*. Birkhäuser.

Manyi-Loh, C., Mamphweli, S., Meyer, E., & Okoh, A. (2018). Antibiotic use in agriculture and its consequential resistance in environmental sources: Potential public health implications. *Molecules*, *23*(4), 795. https://doi.org/10.3390/molecules23040795

Markov, A. A. (1954). *Theory of algorithms* [Translated by Jacques J. Schorr-Kon and PST staff. Imprint Moscow, Academy of Sciences of the USSR, 1954. Jerusalem, Israel Program for Scientific Translations, 1961; available from Office of Technical Services, United States Department of Commerce. Added t.p. in Russian. Translation of Works of the Mathematical Institute, Academy of Sciences of the USSR, v. 42. Original title: Teoriya algorifmov. [QA248.M2943 Dartmouth College library. U.S. Dept. of Commerce, Office of Technical Services, number OTS 60-51085.]]. Academy of Sciences of the USSR.

Markowiak, P., & Śliżewska, K. (2018). The role of probiotics, prebiotics and synbiotics in animal nutrition. *Gut pathogens*, *10*(1), 1–20.

Muthén, B., & Asparouhov, T. (n.d.). Transition probabilities influenced by covariates. *Mplus Web Notes*, (13).

Pérez-Ocón, R., Ruiz-Castro, J. E., & Gámiz-Pérez, M. L. (n.d.). Semi-markov models for lifetime data analysis.

Putter, H., Fiocco, M., & Geskus, R. B. (2007a). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, *26*(11), 2389–2430.

Putter, H., Fiocco, M., & Geskus, R. B. (2007b). Tutorial in biostatistics: Competing risks and multi-state models [Citations: 1,439]. *Statistics in Medicine*, *26*(11), 2389–2430. https://doi.org/10.1002/sim.2712

Romero, J., Benavides, E., & Meza, C. (2018). Assessing financial impacts of subclinical mastitis on colombian dairy farms. *Frontiers in Veterinary Science*, *5*, 273. https://doi.org/10.3389/fvets.2018.00273

Rondeau, V., Mazroui, Y., & Gonzalez, J. (2003). 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*.

Ruet, P., Maincent, P., Berthelot, X., & Kaltsatos, V. (2001). Bovine mastitis and intramammary drug delivery: Review and perspectives. *Advanced Drug Delivery Reviews*, *50*, 245–259. https://doi.org/10.1016/S0169-409X(01)00160-0

Schmid, M., Küchenhoff, H., Hoerauf, A., & Tutz, G. (2016). A survival tree method for the analysis of discrete event times in clinical and epidemiological studies. *Statistics in Medicine*, *35*(5), 734–751. https://doi.org/10.1002/sim.6792

Sharma. (2011). Antibiotic resistance in ocular bacterial pathogens. *Journal of Medical Microbiology*, *60*(5), 535–542.

Sharma, N., Singh, N., & Bhadwal, M. (2011). Relationship of somatic cell count and mastitis: An overview. *Asian-Australasian Journal of Animal Sciences*, *24*, 429–438. https://doi.org/10.5713/ajas.2011.10233

Sharun, K., Dhama, K., Tiwari, R., Gugjoo, M. B., Yatoo, M. I., Patel, S. K., Pathak, M., Karthik, K., Khurana, S. K., Singh, R., Puvvala, B., Amarpal, Singh, R., Singh, K. P., & Chaicumpai, W. (2021). Advances in therapeutic and managemental approaches of bovine mastitis: A comprehensive review [Published online 2021 Feb 17]. *Veterinary Quarterly*, *41*(1), 107–136. https://doi.org/10.1080/01652176.2021.1882713

Springer. (2008). Unobserved heterogeneity. In *Econometric analysis of count data*. https://doi.org/10.1007/978-3-540-78389-3_4

Tapak, L., Kosorok, M. R., Sadeghifar, M., & Hamidi, O. (2018). Multistate recursively imputed survival trees for time-to-event data analysis: An application to aids and mortality post-hiv infection data. *BMC Medical Research Methodology*, *18*. https://doi .org/10.1186/s12874-018-0571-7

Tomanić, D., Samardžija, M., & Kovačević, Z. (2023). Alternatives to antimicrobial treatment in bovine mastitis therapy: A review [Published online 2023 Mar 30]. *Antibiotics (Basel)*, *12*(4), 683. https://doi.org/10.3390/antibiotics12040683

Van Belle, V., Pelckmans, K., Suykens, J. A., & Van Huffel, S. (2008). A practical scalable algorithm. https://doi.org/10.13140/RG.2.1.2920.3681

van Kampen, N. (1998). Remarks on non-markov processes. *Brazilian Journal of Physics*, *28*(2), -. https://doi.org/10.1590/S0103-97331998000200003

Vaupel, J., Manton, K., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*.

Waller, K. P., Persson, Y., Nyman, A.-K., & Stengärde, L. (2014). Udder health in beef cows and its association with calf growth. *Acta Veterinaria Scandinavica*, *56*, 9. https://doi.org/10.1186/1751-0147-56-9

Wan, L., Lou, W., Abner, E., & Kryscio, R. J. (2017). A comparison of time-homogeneous markov chain and markov process multi-state models. *Statistics in Biosciences*, *9*, 92–100. https://doi.org/10.1080/23737484.2017.1361366

Wiener, J. M., & Tilly, J. (2002). Population ageing in the united states of america: Implications for public programmes. *International Journal of Epidemiology*, *31*(4), 776–781.

Yingkui, G., & Jing, L. (2011). Multi-state system reliability: A new and systematic review. *Procedia Engineering*, *11*, 597–603. https://doi.org/10.1016/j.proeng.2011.12.756

Zhao, X., & Lacasse, P. (2008). Mammary tissue damage during bovine mastitis: Causes and control. *Journal of Animal Science*, *86*, 57–65. https://doi.org/10.2527/jas.2007-0302

Zhu, R., & Kosorok, M. R. (2012). Recursively imputed survival trees. *Journal of the American Statistical Association*, *107*(497), 331–340. https://doi.org/10.1080/01621459.2011.645777

Zhu, X., Chen, W.-J., Bhatt, K., Zhou, Z., Huang, Y., Zhang, L.-H., Chen, S., & Wang, J. (2022). Innovative microbial disease biocontrol strategies mediated by quorum quenching and their multifaceted applications: A review [Section: Plant Pathogen Interactions]. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.1063393

# A Treatments

Table 10: Medicine vs Pathogens

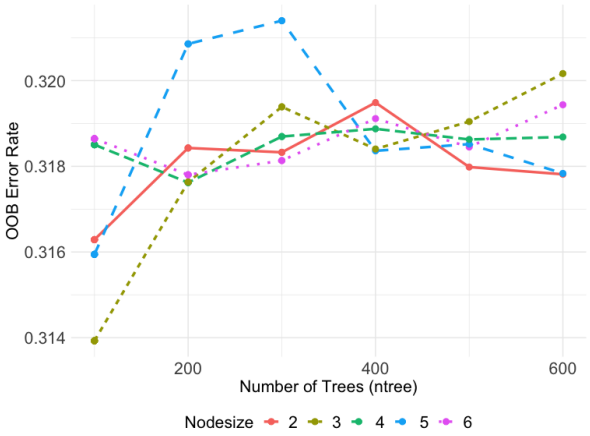
| Medicine name: | For pathogens: | | | |
|---|---|---|---|---|
| | Streptococci: | Staphylococci: | Escherichia coli: | Other: |
| Albiotic formula | x | x | x | – |
| Amphoprim | x | x | x | Klebsiella |
| Ampicillan | – | – | – | Enterobacteriaceae |
| Avuloxil | x | x | x | – |
| Diatrim | x | x | x | Klebsiella |
| Dofatrim | x | x | x | Klebsiella |
| Mamyzin | x | x | – | – |
| Orbenin lactation | x | x | – | – |
| Tylucyl | x | x | – | – |
| Ubrolexin | x | x | x | – |
| Ubropen | x | x | – | — |

# B Numerical overview results

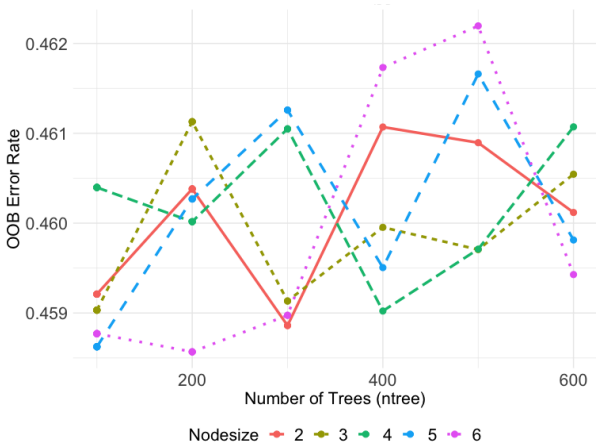
Table 11: C-Index, likelihood and IBS score for different company sizes

| Sample Size | Condition | Metric | Healthy to Sick | Sick to Healthy | Sick to Died | Healthy to Died |
|---|---|---|---|---|---|---|
| n = 25 | normal | concordance | 0.625 | 0.618 | 0.743 | 0.782 |
| n = 25 | normal | likelihood | 11.42 | 14.668 | 12.524 | 17.104 |
| n = 25 | normal | IBS | 0.215 | 0.104 | 0.079 | 0.411 |
| n = 25 | frailty | concordance | 0.769 | 0.768 | 0.767 | 0.807 |
| n = 25 | frailty | likelihood | 57.52 | 74.2 | 12.604 | 17.108 |
| n = 25 | frailty | IBS | 0.215 | 0.107 | 0.079 | 0.411 |
| n = 25 | MSRIST | concordance | 0.53 | 0.601 | 0.7169 | 0.737 |
| n = 25 | MSRIST | IBS | 0.153 | 0.102 | 0.024 | 0.081 |
| n = 50 | normal | concordance | 0.59 | 0.576 | 0.735 | 0.789 |
| n = 50 | normal | likelihood | 5.024 | 5.08 | 7.954 | 12.482 |
| n = 50 | normal | IBS | 0.216 | 0.187 | 0.155 | 0.267 |
| n = 50 | frailty | concordance | 0.776 | 0.778 | 0.761 | 0.811 |
| n = 50 | frailty | likelihood | 42.9 | 57.4 | 7.986 | 12.844 |
| n = 50 | frailty | IBS | 0.212 | 0.188 | 0.155 | 0.267 |
| n = 50 | MSRIST | concordance | 0.571 | 0.56 | 0.6778 | 0.741 |
| n = 50 | MSRIST | IBS | 0.155 | 0.099 | 0.0577 | 0.15665 |
| n = 100 | normal | concordance | 0.575 | 0.56 | 0.755 | 0.791 |
| n = 100 | normal | likelihood | 2.996 | 2.693 | 6.029 | 8.519 |
| n = 100 | normal | IBS | 0.214 | 0.178 | 0.185 | 0.254 |
| n = 100 | frailty | concordance | 0.779 | 0.792 | 0.782 | 0.813 |
| n = 100 | frailty | likelihood | 31.92 | 40.51 | 6.09 | 8.52 |
| n = 100 | frailty | IBS | 0.211 | 0.179 | 0.185 | 0.254 |
| n = 100 | MSRIST | concordance | 0.566 | 0.548 | 0.704 | 0.737 |
| n = 100 | MSRIST | IBS | 0.152 | 0.081 | 0.076 | 0.1475 |
| n = 250 | normal | concordance | 0.56 | 0.543 | 0.748 | 0.768 |
| n = 250 | normal | likelihood | 1.3964 | 1.0824 | 3.0144 | 4.056 |
| n = 250 | normal | IBS | 0.217 | 0.165 | 0.196 | 0.274 |
| n = 250 | frailty | concordance | 0.778 | 0.794 | 0.782 | 0.795 |
| n = 250 | frailty | likelihood | 18.016 | 21.736 | 3.9628 | 4.056 |
| n = 250 | frailty | IBS | 0.214 | 0.165 | 0.198 | 0.274 |
| n = 250 | MSRIST | concordance | 0.551 | 0.544 | 0.709 | 0.718 |
| n = 250 | MSRIST | IBS | 0.152 | 0.074 | 0.08 | 0.148 |
| n = 500 | normal | concordance | 0.554 | 0.537 | 0.736 | 0.751 |
| n = 500 | normal | likelihood | 0.7422 | 0.5564 | 1.619 | 2.138 |
| n = 500 | normal | IBS | 0.219 | 0.154 | 0.2 | 0.276 |
| n = 500 | frailty | concordance | 0.773 | 0.79 | 0.838 | 0.781 |
| n = 500 | frailty | likelihood | 10.052 | 11.724 | 3.326 | 2.156 |
| n = 500 | frailty | IBS | 0.217 | 0.154 | 0.209 | 0.276 |
| n = 500 | MSRIST | concordance | 0.555 | 0.544 | 0.704 | 0.717 |
| n = 500 | MSRIST | IBS | 0.152 | 0.075 | 0.078 | 0.147 |
| n = 1000 | normal | concordance | 0.549 | 0.554 | 0.735 | 0.749 |
| n = 1000 | normal | likelihood | 0.3672 | 0.2741 | 0.838 | 1.088 |
| n = 1000 | normal | IBS | 0.22 | 0.152 | 0.205 | 0.284 |
| n = 1000 | frailty | concordance | 0.768 | 0.791 | 0.842 | 0.78 |
| n = 1000 | frailty | likelihood | 4.94 | 6.025 | 1.85 | 1.107 |
| n = 1000 | frailty | IBS | 0.219 | 0.153 | 0.216 | 0.284 |
| n = 1000 | MSRIST | concordance | 0.541 | 0.523 | 0.71 | 0.688 |
| n = 1000 | MSRIST | IBS | 0.152 | 0.068 | 0.079 | 0.147 |

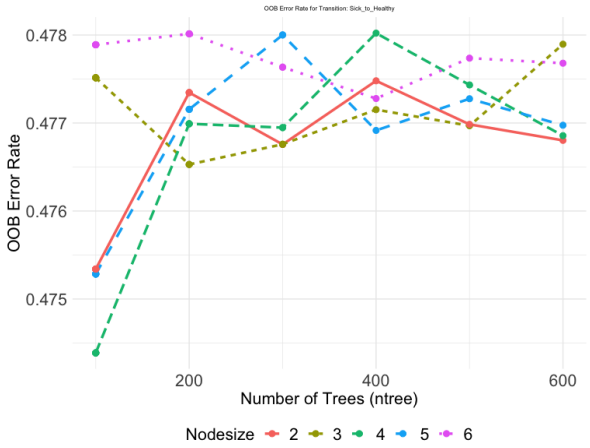
# C Results Parameter Optimalisation



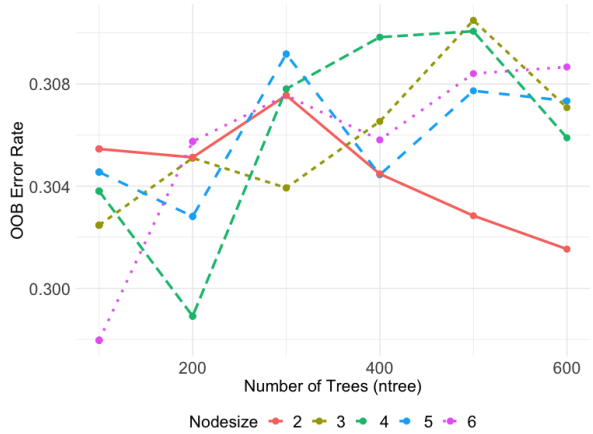

(a) Optimalisation for Healthy to Died




(b) Optimalisation for Healthy to Sick




(c) Optimalisation for Sick to Healthy




(d) Optimalisation for Sick to Died

Figure 10: Parameter Optimalisation for MSRIST