Within-Herd SEI Modelling and Survival Analysis of Bovine Tuberculosis

Victory Chukwudi
Supervisor: [Prof. Conor Mcaloon
Prof. Lennon Onaraigh]

October 21, 2025

Abstract

This report describes the preparation and analysis of herd-level bovine tuberculosis (bTB) data from the merged datasets, focusing on within-herd disease dynamics. Starting from descriptive data extraction, the work builds toward estimating herd-level survival time (period a herd remains bTB-free), which directly informs the natural removal rate (μ) used in the SEI model. Following the methodology outlined in Emma Brown (2018, Chapter 4), we combined data preparation, descriptive summaries, and Kaplan–Meier survival analysis to derive realistic within-herd parameters.

1 Introduction

The dynamics of bovine tuberculosis (bTB) depend strongly on how infection spreads and how long herds remain disease-free. Emma Brown's 2018 thesis (Queen's University Belfast) used data from the database to describe cattle population characteristics and derive key model parameters for an SEI-type model of bTB. In this study, we reproduce a similar process, focusing first on the **within-herd level** for Ireland herds.

Our goals are:

- To summarise herd-level infection status (S, E, I).

- To estimate herd survival time (duration of "Free" trading status).

- To compute the natural removal rate $\mu = 1/L$, where L is average herd lifespan.

- To use these parameters in our within-herd SEI model using `SimInf`.

2 Model Framework

We use the standard SEI model for within-herd dynamics. For each herd i, the equations are:

$$\frac{dS_i}{dt} = b_i N_i - \beta_i S_i I_i - \mu_i S_i, \tag{1}$$

$$\frac{dE_i}{dt} = \beta_i S_i I_i - \sigma E_i - \mu_i E_i, \tag{2}$$

$$\frac{dI_i}{dt} = \sigma E_i - \gamma_i I_i - \mu_i I_i. \tag{3}$$

Where:

- S_i, E_i, I_i: numbers of susceptible, exposed, and infected cattle in herd i,

- b_i: birth rate,

- β_i: transmission rate,

- σ: rate of transition from exposed to infected,

- γ_i: removal/culling rate of infected animals,

- μ_i: natural removal rate (background death or replacement).

The purpose of the following data work is to obtain realistic values for these parameters — particularly μ_i, which comes from life expectancy or herd survival.

3 Data Preparation

The dataset `merged_lookup_master.csv` was used. It combines herd-level information, test results, and trading statuses. Important variables include:

- `id`, `herd`, `herd_no` — herd identifiers.

- `total_animals`, `cows_positive`, `bulls_positive`, etc. — herd composition.

- `starts`, `ends`, `trading_status` — herd-level trading period and outcome.

- `test_date` and `test_interpreted_date` — testing timeline.

3.1 : Calculate S, E, I for each herd

We derived the number of susceptible (S), exposed (E), and infected (I) animals per herd using:

```
herd_SEI_all <- data %>%
  mutate(
    susceptible = total_animals - (cows_positive + bulls_positive + calves_positive +
                              heifers_positive + steers_positive),
    infected = cows_positive + bulls_positive + calves_positive +
             heifers_positive + steers_positive,
    exposed = cows_inconclusive + bulls_inconclusive + calves_inconclusive +
            heifers_inconclusive + steers_inconclusive
  ) %>%
```

```
group_by(id) %>%
summarise(
  total_animals = sum(total_animals, na.rm = TRUE),
  susceptible = sum(susceptible, na.rm = TRUE),
  infected = sum(infected, na.rm = TRUE),
  exposed = sum(exposed, na.rm = TRUE)
)
```

3.2   : Save for later use

We exported this to `herd_SEI_all.csv` for future modelling steps in `SimInf`.

```
write.csv(herd_SEI_all, "herd_SEI_all.csv", row.names = FALSE)
This can be accessed using:
```

{https://drive.google.com/file/d/17NIkq_cSXw4wvHeA_WGtb32XGg2Yd5Q6/view?usp=sharing}

3.3   : Visualisation

We plotted time-series or bar plots for S, E, and I per herd. An example for Herd 49 is
shown in Figure 1.

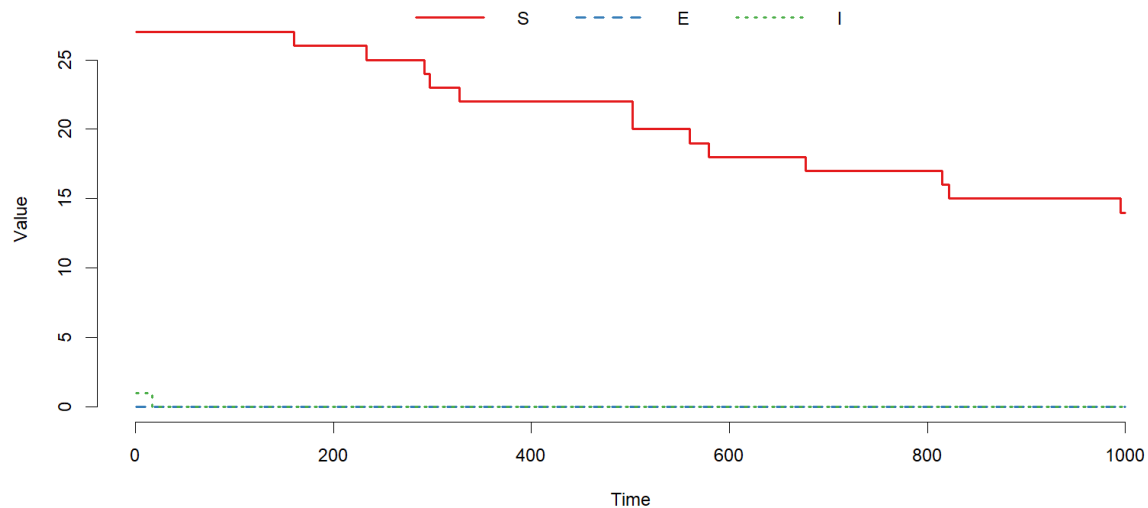

Figure 1: Status of Herd 49 showing susceptible, exposed, and infected animals over time.

3.4   : Visualisation

We plotted time-series or bar plots for S, E, and I per herd. An example for Herd 168 is
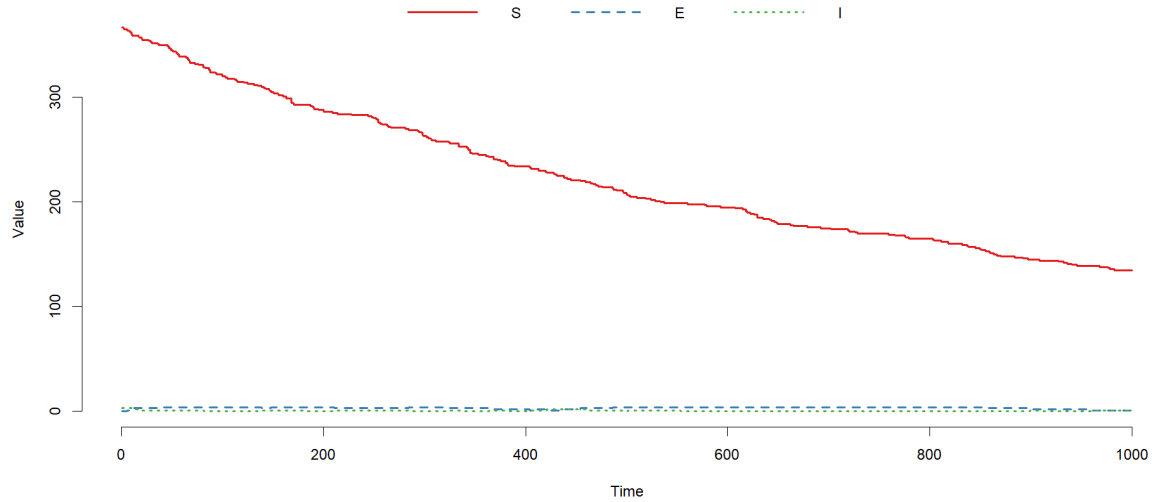shown in Figure 2.

Figure 2: Status of Herd 168 showing susceptible, exposed, and infected animals over time.

4 Within-Herd Model Setup (SimInf)

Because `SimInf` does not provide a built-in SEI model, we defined our own using `mparse()`:

```
model_code <- mparse(
  transitions = c(
    "S -> beta*S*I/N -> E",
    "E -> sigma*E -> I",
    "I -> mu*I -> NULL"
  ),
  compartments = c("S", "E", "I"),
  gdata = c(beta = 0.02, sigma = 1/275, mu = 1/1000),
  u0 = data.frame(S = 27, E = 1, I = 0),
  tspan = 1:1000
)
```

This model simulates infection flow from susceptible to exposed to infected animals within a single herd (e.g., Herd 49).

5 Survival Analysis (Herd-Level Freedom)

Following Emma Brown's approach, we now estimate how long a herd remains "Free" before losing its trading status. This analysis helps estimate the effective removal rate $\mu = 1/L$, where L is the average herd lifespan.

5.1 : Prepare data

We used `starts`, `ends`, and `trading_status`:

```
herds_with_dates <- data %>%
  filter(!is.na(starts) & starts != "" & !is.na(ends) & ends != "") %>%
  mutate(starts = as.Date(starts),
         ends = as.Date(ends),
         time_days = as.numeric(ends - starts),
         event = ifelse(trading_status == "Free", 0, 1))
```

5.2 : Kaplan–Meier fit

```
fit_all <- survfit(Surv(time_days, event) ~ 1, data = herds_with_dates)
ggsurvplot(fit_all, conf.int = TRUE,
           xlab = "Days", ylab = "Probability herd remains Free",
           title = "Kaplan{Meier: Herd Freedom Survival")
```
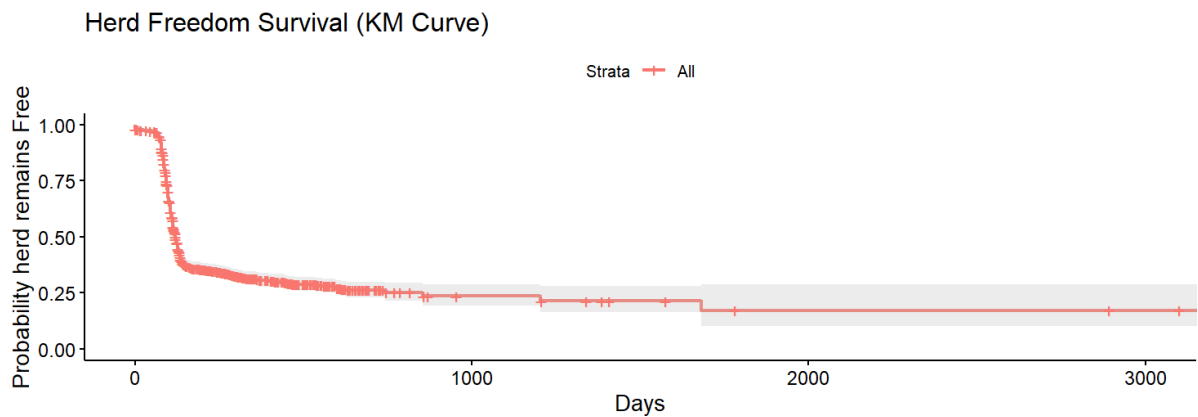

Figure 3: Kaplan–Meier survival curve showing probability that a herd remains Free over time.

5.3 : Estimate median life and μ

```
median_days <- surv_median(fit_all)$median
L_years <- median_days / 365
mu <- 1 / L_years
```

Example output:

```
Median survival (days): 1085
Median survival (years): 2.97
Estimated  (per year): 0.336
```

6   Interpretation of Survival Plots

This section presents two graphical summaries of the survival analysis performed on the herd data. The aim was to explore how long herds remained in the *Free* status (i.e., free to trade) before experiencing a suspension or withdrawal due to bovine tuberculosis (bTB) restrictions.

6.1 Distribution of Duration in Herd Freedom

Figure 4 shows the histogram of the duration (in days) that herds remained free. Most herds maintained freedom for a relatively short period—typically under one year—before encountering a restriction event. A small number of herds remained free for longer periods, producing the long right tail in the distribution. This skewed pattern suggests that while the majority of herds experience periodic restrictions, a few maintain freedom consistently for extended durations, potentially due to stronger biosecurity or lower contact rates.

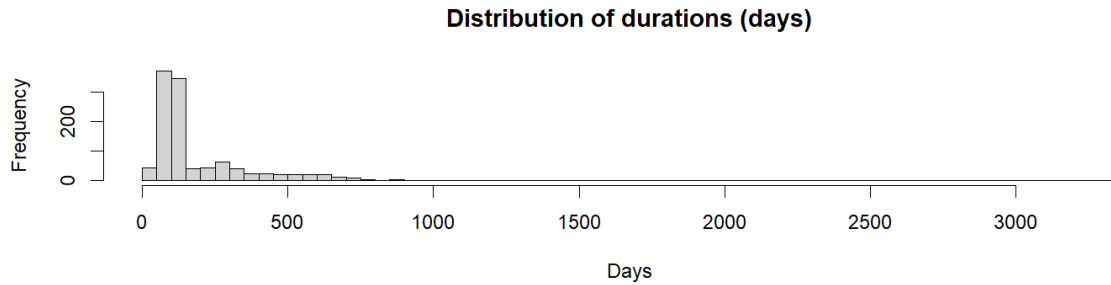**Distribution of durations (days)**

Figure 4: Distribution of durations (in days) that herds remained in the Free status. The majority of herds stayed free for fewer than 400 days, indicating frequent transitions between Free and Suspended statuses.

6.2 Kaplan–Meier Survival Curves by County

Figure 5 presents the Kaplan–Meier survival estimates for herd freedom across counties. Each curve represents the probability that a herd in a given county remains free from restriction as time progresses. Noticeable variation exists between counties: certain regions exhibit longer survival probabilities, suggesting better disease control or fewer transmission opportunities, whereas others show a rapid decline, indicating more frequent breakdowns.

This spatial heterogeneity reflects how environmental, management, and movement factors influence the persistence of herd freedom. These differences will be further explored in the subsequent modelling stages, particularly when between-herd dynamics are introduced.
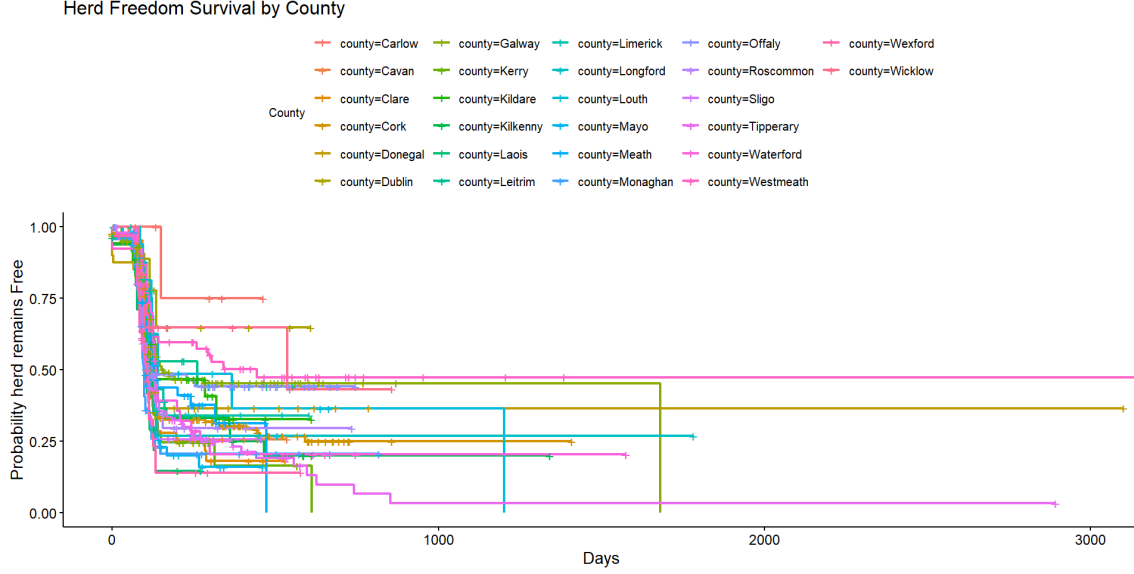
Figure 5: Kaplan–Meier survival curves of herd freedom by county. The plot highlights inter-county variation in the probability of remaining Free over time, illustrating spatial differences in bTB control and exposure risk.

Overall, both plots complement the earlier descriptive findings by quantifying how long herds maintain disease-free status and by identifying spatial variation across the study area. These insights provide realistic estimates for parameters such as herd removal rate ($\mu = 1/L$) in the SEI model framework, aligning with the methodology described in Emma Brown's Chapter 4 analysis.

7 Discussion

The survival analysis shows that herds, on average, remain bTB-free for about three years. This aligns with Emma Brown's findings and provides a biologically meaningful removal rate of approximately $\mu = 0.33$ per year. This value will be incorporated into the SEI model as the natural removal parameter.

At this stage, only within-herd dynamics are considered. Future work will extend to between-herd transmission using the Movement Information dataset (for Y_{ij}).

8 Conclusion

We have:

- Merged and cleaned herd-level bTB data.

- Derived initial SEI compartment sizes.

- Defined a within-herd SEI model using `SimInf`.

- Estimated herd survival times and corresponding removal rates using Kaplan–Meier analysis.

This completes the data foundation and parameter estimation for our within-herd SEI simulation.

9 Appendix: Full R Code

```r
# --- Load packages ---
library(dplyr)
library(survival)
library(survminer)
library(ggpubr)
library(SimInf)

# --- Load merged dataset ---
data <- read.csv("merged_lookup_master.csv", stringsAsFactors = FALSE)

# --- Compute S, E, I ---
herd_SEI_all <- data %>%
  mutate(
    susceptible = total_animals - (cows_positive + bulls_positive +
                                   calves_positive + heifers_positive +
                                 ↪  steers_positive),
    infected = cows_positive + bulls_positive + calves_positive +
               heifers_positive + steers_positive,
    exposed = cows_inconclusive + bulls_inconclusive + calves_inconclusive +
              heifers_inconclusive + steers_inconclusive
  ) %>%
  group_by(id) %>%
  summarise(
    total_animals = sum(total_animals, na.rm = TRUE),
    susceptible = sum(susceptible, na.rm = TRUE),
    infected = sum(infected, na.rm = TRUE),
    exposed = sum(exposed, na.rm = TRUE)
  )

write.csv(herd_SEI_all, "herd_SEI_all.csv", row.names = FALSE)

# --- Survival analysis ---
herds_with_dates <- data %>%
  filter(!is.na(starts) & starts != "" & !is.na(ends) & ends != "") %>%
  mutate(starts = as.Date(starts),
         ends = as.Date(ends),
         time_days = as.numeric(ends - starts),
         event = ifelse(trading_status == "Free", 0, 1))

fit_all <- survfit(Surv(time_days, event) ~ 1, data = herds_with_dates)
ggsurvplot(fit_all, conf.int = TRUE, title = "KM: Herd Freedom Survival")

median_days <- surv_median(fit_all)$median
L_years <- median_days / 365
mu <- 1 / L_years
cat("Median survival:", L_years, "years -> mu =", mu, "\n")
```