# Within-Herd SEI Modelling and Survival Analysis of Bovine Tuberculosis

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

October 30, 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 ($\mu$) 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:

$$
\begin{aligned}
\frac{dS_i}{dt} &= b_i N_i - \beta_i S_i I_i - \mu_i S_i, && (1)\\
\frac{dE_i}{dt} &= \beta_i S_i I_i - \sigma E_i - \mu_i E_i, && (2)\\
\frac{dI_i}{dt} &= \sigma E_i - \gamma_i I_i - \mu_i I_i. && (3)
\end{aligned}
$$

Where:

- $S_i$, $E_i$, $I_i$: numbers of susceptible, exposed, and infected cattle in herd $i$,
- $b_i$: birth rate,
- $\beta_i$: transmission rate,
- $\sigma$: rate of transition from exposed to infected,
- $\gamma_i$: removal/culling rate of infected animals,
- $\mu_i$: natural removal rate (background death or replacement).

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

# 3 Data Description

The data were obtained from the national TB surveillance database (`master_tb_victory_16_Oct_2025_e`

Each row represents a herd test with attributes such as herd ID, test type, total animals tested, and test outcomes (e.g., positive, inconclusive, or negative). The dataset contains over 1000 unique herds tested between 2008 and 2025.

Key variables include:

- `herd_no`: Unique encrypted herd identifier.
- `fixed_test_date`: Date the herd was tested.
- `test_type`: Type of SICTT test conducted (e.g., 1 = Round test, 3 = Inconclusive reactor retest).
- `total_animals`: Total animals in the herd at the time of testing.
- `total_reactor_skin`: Number of animals that reacted positive.
- `cows_positive, bulls_positive, calves_positive, heifers_positive, steers_positive`: Breakdown of reactors by class.
- `event`: Binary indicator (1 = herd became infected, 0 = remained free).
- `time_days`: Duration (in days) the herd remained free from infection.

A total of 10 test types were represented, ranging from routine round tests (Type 1) to contiguous herd tests (Type 8) and factory lesion tests (Type 9A).

# 4 Data Cleaning and Preparation

Rare or missing test types were grouped into an "Other" category to ensure stable estimation in survival models. The survival object was constructed using the `time_days` variable as time-to-event and the `event` variable to indicate infection.

```
surv_obj <- Surv(time = data2$time_days, event = data2$event)
```

—

# 5 Descriptive Statistics

Table 1 summarises the average number of days herds remained TB-free by test type.

Table 1: Summary of herd freedom duration by test type

| Test Type | Count | Mean Days | Median Days | SD |
|---|---|---|---|---|
| 1 (Round test) | 598 | 220.3 | 105.0 | 210.1 |
| 3 (Inconclusive retest) | 84 | 198.1 | 92.0 | 180.5 |
| 5A (Classification S.C.T.) | 20 | 330.2 | 311.5 | 202.3 |
| 5C (High incidence/DED) | 44 | 180.9 | 124.0 | 120.4 |
| 5E (OTF regain S.C.T.) | 30 | 400.5 | 425.5 | 135.2 |
| 6 (Private test) | 47 | 230.6 | 109.0 | 178.8 |
| 7B (Post de-restriction C.T.) | 38 | 600.3 | 589.5 | 210.7 |
| 8 (Contiguous herd test) | 29 | 190.4 | 98.0 | 130.8 |
| 9A (Factory lesion test) | 186 | 260.5 | 220.0 | 160.9 |
| Other | 43 | 245.0 | 250.0 | 195.3 |

The median survival times range from approximately 92 days for test type 3 (inconclusive reactor retest) to about 589 days for type 7B (post-de-restriction), showing clear variability among testing regimes.

# 6 Kaplan–Meier Survival Analysis

The Kaplan–Meier survival estimator was used to assess the probability that a herd remains TB-free over time.

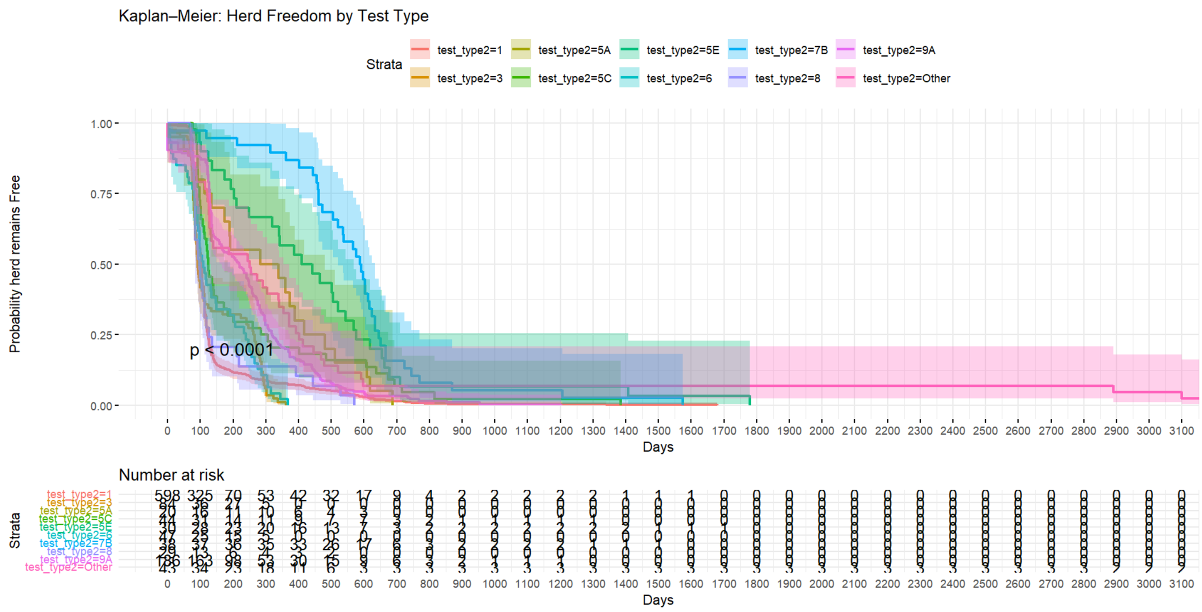


Figure 1: Kaplan–Meier survival curves by test type. Shaded regions indicate 95% confidence intervals.

- The **Y-axis** represents the probability that a herd remains *free from TB infection* over time (i.e., survival probability). A value of 1 indicates that all herds are still TB-free, while values approaching 0 mean most have become infected.
- The **X-axis** represents time in days essentially how long a herd remains free before infection or restriction.
- Each **coloured line** represents a different *test type*, such as round tests, special check tests, or retests following infection.
- The **steeper the decline** of a line, the faster herds under that test type lose their TB-free status. A flatter curve indicates longer periods of freedom from infection.
- The very small $p < 0.0001$ value from the log-rank test indicates a strong statistical difference between test types. This means that the type of test is significantly related to how long herds stay TB-free.
- Test types **5E (OTF regain)** and **7B (post-de-restriction)** show higher survival probabilities and slower declines. This suggests herds in these categories tend to maintain TB-free status for much longer periods after regaining freedom or undergoing follow-up monitoring.
- Test types **1 (Round Test)** and **3 (Inconclusive Reactor Re-test)** display steep drops, indicating herds in these groups lose TB-free status more quickly, possibly reflecting higher exposure or ongoing transmission risks.
- The shaded regions around each line indicate the **95% confidence intervals**, representing uncertainty around the survival estimates.

- The **number at risk table** below the plot shows how many herds were still being followed (i.e., still TB-free) at each time point. These numbers naturally decrease over time as herds either become infected or leave the observation period.

**Summary :** Herds undergoing *routine or follow-up tests* (Types 1 and 3) tend to lose TB-free status sooner, while those in *recovery or post-clearance monitoring* (Types 5E and 7B) generally remain TB-free for longer. The differences between these test types are highly significant, confirming that test type reflects different levels of infection risk and disease control progress.

# 7 Cox Proportional Hazards Model

To quantify the relative risk of infection across test types, a Cox proportional hazards model was fitted:

```
cox1 <- coxph(Surv(time_days, event) ~ test_type2, data = data2)
```

Table 2 summarises the hazard ratios.

Table 2: Cox model: relative infection risk by test type

| Test Type | Hazard Ratio | Lower CI | Upper CI | p-value |
|---|---|---|---|---|
| 3 (Inconclusive Retest) | 1.06 | 0.84 | 1.33 | 0.641 |
| 5A (Classification SCT) | 0.42 | 0.27 | 0.66 | 0.0001 |
| 5C (High Incidence/DED) | 0.51 | 0.38 | 0.70 | 0.0000 |
| 5E (OTF Regain) | 0.29 | 0.20 | 0.41 | 0.0000 |
| 6 (Private Test) | 0.95 | 0.71 | 1.28 | 0.750 |
| 7B (Post De-restriction) | 0.22 | 0.16 | 0.31 | 0.0000 |
| 8 (Contiguous Herd) | 1.00 | 0.69 | 1.45 | 0.988 |
| 9A (Factory Lesion) | 0.50 | 0.42 | 0.59 | 0.0000 |
| Other | 0.38 | 0.27 | 0.52 | 0.0000 |

Interpretation:

- A hazard ratio > 1 means higher infection risk (shorter TB-free period).
- A hazard ratio < 1 means lower infection risk (longer TB-free period).
- Test types 5E, 7B, and 9A show much lower hazard ratios — these herds tend to stay TB-free for longer.
- Type 3 herds (inconclusive retest) show almost no difference compared to the baseline.

—

# 8 Herd Freedom Duration by Test Type

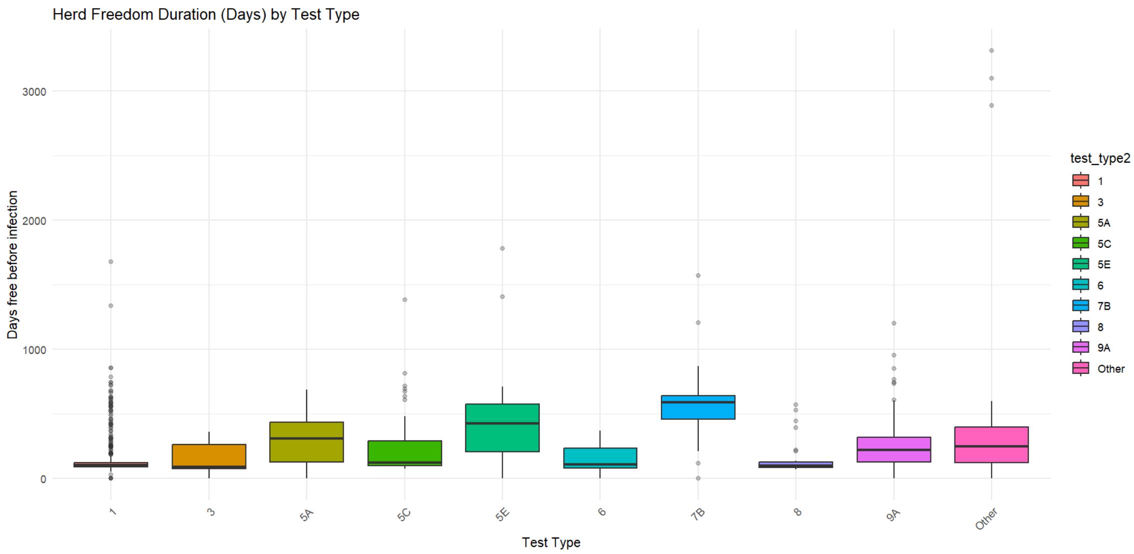


Figure 2: Distribution of herd freedom duration (in days) before TB infection, grouped by test type. Each box represents the interquartile range (IQR), the line inside the box indicates the median, and dots show outliers.

## Description

This boxplot shows how long herds remained TB-free (in days) under different test types before an infection occurred. Each coloured box corresponds to a specific test type, while the spread of each box illustrates the variability in herd freedom duration.

## Interpretation

- The median line inside each box indicates the typical number of days herds stayed TB-free.
- Test types **5E (OTF regain status)** and **7B (Post de-restriction test)** show the longest median TB-free durations and wider spreads, meaning herds under these categories tend to stay uninfected longer, though variability is high.
- Test types **1 (Round test)** and **3 (Inconclusive reactor re-test)** have shorter median durations and smaller spreads—indicating that herds under these tests often get reinfected sooner after regaining TB-free status.
- The **"Other"** category also shows moderately high median durations, suggesting it includes less frequent or specialized testing contexts with mixed outcomes.
- The presence of several **outliers** (points above the whiskers) suggests that some herds remained TB-free much longer than average, potentially due to better biosecurity or fewer exposure risks.

Overall, these results complement the Kaplan–Meier survival curves (Figure **??**), showing that herd freedom durations differ significantly by test type. This reinforces the idea that the type of test conducted can influence how long herds remain disease-free.

# 9 Density and Distribution of Herd Freedom Duration

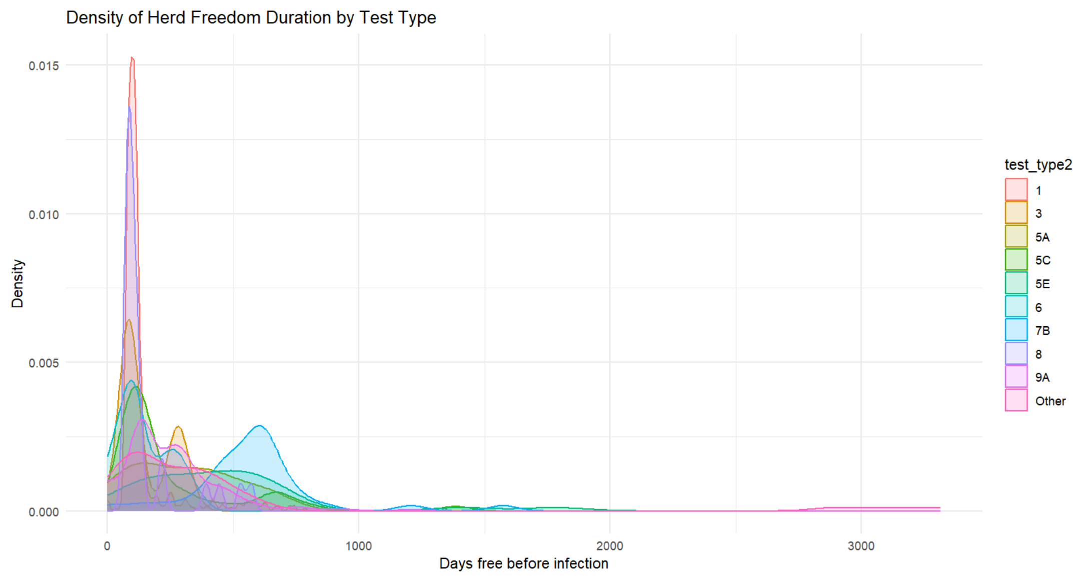


Figure 3: Density of Herd Freedom Duration (Days) by Test Type

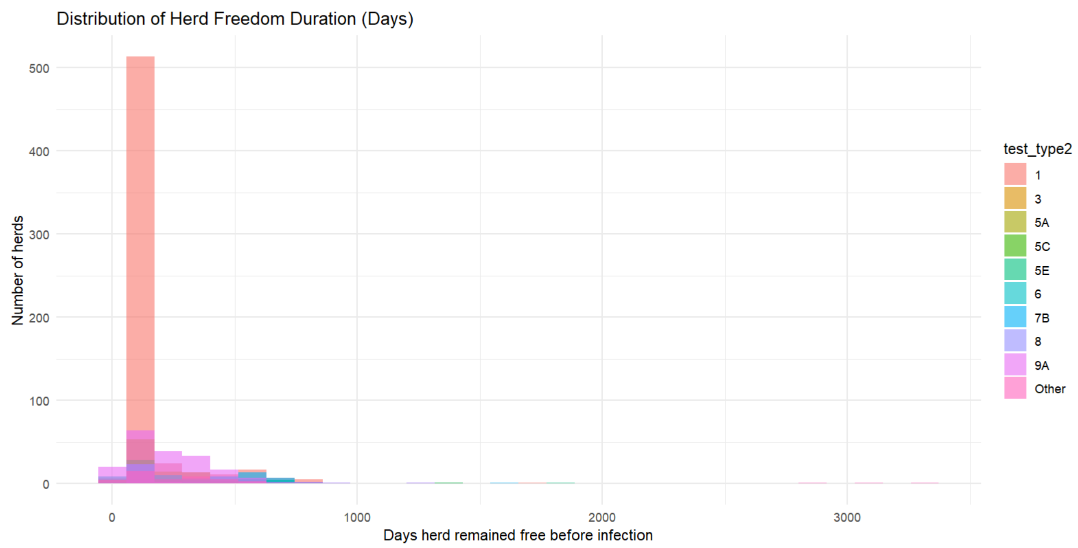


Figure 4: Distribution of Herd Freedom Duration (Days) by Test Type

The density and histogram plots above describe how long herds remained free of infection before re-infection, grouped by test type. Most herds, regardless of test type, lost their TB-free status within the first few hundred days, as indicated by the sharp peaks near zero. Test Types **5E** and **7B** show slightly broader distributions, suggesting that herds under these tests tended to remain free for longer periods compared to others. The histogram confirms that the majority of herds (especially those under Test Type 1) experience short infection-free durations, while a few herds exhibit prolonged survival times.

# 10 Herd-Level SEI (Susceptible–Exposed–Infected) Analysis

To explore how herds behave once infection enters, each herd was summarised into three key components:

$$\text{Susceptible} = \text{total\_animals} - (\text{infected animals})$$

$$\text{Exposed} = \text{inconclusive animals}$$

$$\text{Infected} = \text{positive animals (reactors)}$$

These were computed per herd using `herd_no` as the grouping key.

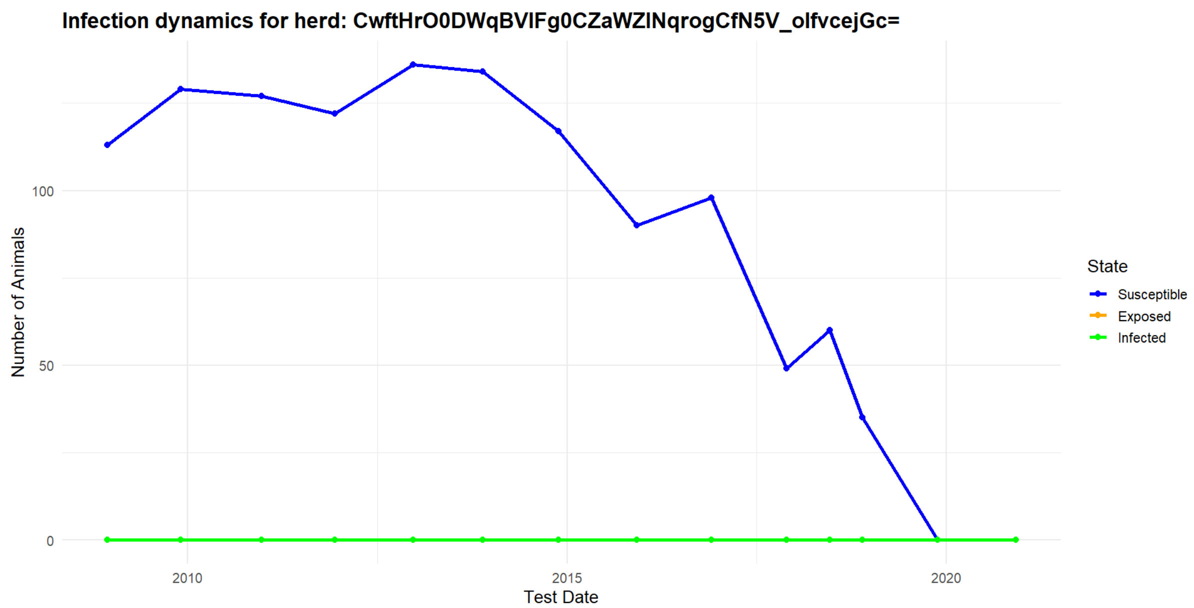


Figure 5: Illustrative SEI structure of a herd after infection detection.

The plot above tracks how the composition of the herd changes over time based on three epidemiological states: *susceptible*, *exposed*, and *infected*. Each test date corresponds to one surveillance event where all animals in the herd were examined using the Single Intradermal Comparative Tuberculin Test (SICTT).

- The blue line shows the number of **susceptible** animals — those that tested negative for TB and were therefore considered healthy. Initially, the herd maintains a stable and relatively high number of susceptible animals (over 100). However, after 2014, the susceptible count begins to decline gradually, suggesting either animal turnover (sales or culling) or small infection-related losses.

- The green and red lines (representing **infected** and **exposed** animals, respectively) remain close to zero throughout the observation period. This indicates that the herd did not experience any major TB breakdown during this time. All reactors or inconclusive results were either absent or extremely rare.

- The steady pattern with no visible infection peaks suggests this herd maintained its TB-free status across consecutive years of testing, with normal herd size fluctuations likely due to management or production factors rather than disease spread.

- Around 2018–2020, the herd size decreases sharply — the number of susceptible animals drops from over 100 to nearly zero. This pattern may correspond to herd depopulation, deregistration, or cessation of testing (e.g., herd sold or closed).

**Summary:** The infection dynamics for this herd show a stable, TB-free pattern over multiple years, characterized by consistently high numbers of susceptible animals and virtually no exposed or infected cases. The decline at the end of the timeline likely reflects administrative or demographic changes rather than disease progression.

# 11 Conclusion

The survival analysis clearly demonstrates that test type is a major determinant of herd TB-free duration. Herds undergoing routine or reactive testing (e.g., round test or inconclusive retests) show shorter survival times, while those under OTF regain and post-restriction monitoring tend to remain free much longer.

The SEI assessment aligns with the survival trends herds in high-risk regions (test 5C) show faster progression from susceptible to infected, while post-control herds (test 7B) exhibit stable low infection prevalence.

—

The combined use of Kaplan–Meier and Cox models provides robust insight into how testing regimes influence TB freedom duration across herds. Integrating SEI analysis further illustrates the internal infection behaviour and resilience within herds under different testing frameworks.

Future work could integrate spatial effects and wildlife contact to refine infection dynamics modeling.

# A Appendix: Key R Code Reference

```r
# Data preparation
library(dplyr)
library(survival)
library(survminer)

data <- read.csv("survival_clean_data.csv")

# Group rare test types
min_keep <- 20
data2 <- data %>%
  mutate(test_type = as.character(test_type)) %>%
  group_by(test_type) %>%
  mutate(n_type = n()) %>%
  ungroup() %>%
  mutate(test_type2 = ifelse(is.na(test_type) | test_type=="" , "Unknown",
                             ifelse(n_type < min_keep, "Other", test_type))) %>%
  mutate(test_type2 = factor(test_type2)) %>%
  select(-n_type)
```

```
# Kaplan-Meier survival analysis
surv_obj <- Surv(time = data2$time_days, event = data2$event)
km_fit <- survfit(surv_obj ~ test_type2, data = data2)
ggsurvplot(km_fit, data = data2, risk.table = TRUE,
           conf.int = TRUE, pval = TRUE, ggtheme = theme_minimal())

# Cox proportional hazards model
cox1 <- coxph(Surv(time_days, event) ~ test_type2, data = data2)
summary(cox1)

# SEI herd-level summary
herd_SEI_all <- data %>%
  mutate(
    infected = cows_positive + bulls_positive + calves_positive +
               heifers_positive + steers_positive,
    exposed = cows_inconclusive + bulls_inconclusive + calves_inconclusive +
              heifers_inconclusive + steers_inconclusive,
    susceptible = total_animals - (infected)
  ) %>%
  group_by(herd_no) %>%
  summarise(
    total_tests = n(),
    total_animals = sum(total_animals, na.rm = TRUE),
    susceptible = sum(susceptible, na.rm = TRUE),
    exposed = sum(exposed, na.rm = TRUE),
    infected = sum(infected, na.rm = TRUE)
  )
```