MM925: Quantitative Risk Analysis – Report

University of Strathclyde

Table of Contents

Chapter 1 – Background & Introduction

Echinococcus Multilocularis

Echinococcus multilocularis is a disease that affects the liver, as a tumour causing abdominal pain and sometimes lesions in the lungs and brain. Dogs are the most common hosts for the adult parasites that reside in the small intestine, eggs are released and passed in faeces where they are ingested by intermediary hosts. (McConnaughey, 2014)

Public Health Concern

It represents a significant public health concern as one of the most dangerous parasitic zoonoses in the Northern Hemisphere. If left untreated, the disease has a mortality rate of over 90% within 10-15 years of diagnosis. (Brunetti & Kern, 2010) Even with modern treatment, the prognosis remains serious, requiring long-term medication and often complex surgery. The growth pattern of the parasite makes it particularly challenging to treat.

Economic Impact

Beyond the direct public health implication, the establishment of E. multilocularis in a country like the UK would have significant economic consequences such as:

- Increased healthcare costs for treating human cases
- Implementation of new veterinary screening problems
- Potential effects on wildlife management
- Costs associated with public health monitoring and control measures

Current Protection Status

The U. K's geographic island status, combined with strict travel regulations has in the past protected it from many diseases such as E. multilocularis. However, rising global international travel, new policies such as the Pet Travel Scheme (PETS), and a potential knock-on effect of the EU Pet Travel Scheme has created new routes for potential introduction. The current requirements for dogs entering the UK include:

- Treatment with praziquantel 24-120 hours before entry
- Documentation of treatment in the pet passport

- Compliance checks at the port of entry

Aims

This report aims to quantify the risk of E. multilocularis introduction into the United Kingdom through dogs returning from two-week holidays in France. Through mathematical modelling and statistical analysis, we evaluated:

- The probability of dog infection during holiday stays
- Owner compliance with treatment requirements
- Treatment timing patterns
- Overall annual risk of parasite introduction
- Sensitivity of risk estimates to key parameters

Understanding these factors is crucial for assessing current control measures and informing policy decisions about pet travel regulations. This analysis will provide evidence-based insights to support public health decision-making and protect the U.K's disease-free status.

Chapter 2 – Methods

The study employed probabilistic and statistical methods to assess the risk of Echinococcus multilocularis introduction into the U.K. via dogs return from France. The analysis was conducted using R version 4.40 with the code shown in the Appendix of this report.

Mathematical Model

The probability of a dog becoming infected with Echinococcus multilocularis during a stay in France was modelled using the below equation:

$$p(t) = \frac{v(1 - e^{-(\mu+v)t})}{v + \mu}$$

Where:

- $p(t)$ – Infection probability
- v – Is the infection pressure/ how quickly dogs become infected (measured in weeks)
- μ – Is the recovery rate/ how quickly infected dogs can recover (measured in weeks)
- t – Duration of exposure in weeks
- e – Mathematical constant

Individual Components and Parameters

The value of the infection pressure (v) is uncertain, with data suggesting a range of 0.00001 to 0.0006 and a most likely value of 0.000187. In public health and risk assessment, one must account for uncertainty as one fixed value could underestimate or overestimate the real risk which is why this value has been given as a range. In this report we modelled this uncertainty using normal and beta distributions.

- $(1 - e^{-(\mu+v)t})$ represents the cumulative exposure effect over time
- The negative exponential term models the decay in susceptibility
- $v + \mu$ normalises the probability based on both infection and recovery rates
- This ensures the probability remains bounded between 0 and 1

Bayesian Analysis

Bayesian analysis is a method of statistical inference, named after English mathematician Thomas Bayes, that allows one to combine prior information about a population parameter with evidence from information contained in a simple to guide the statistical inference process. A prior probability distribution for a parameter of interest is specified first. The evidence is then obtained and combined through an application of Bayes's Theorem to provide a posterior probability distribution for the parameter. (The Editors of Encyclopaedia Britannica, 2024)

Treatment compliance was analysed using Bayesian inference. For the 2010 survey data, we employed a beta (1,1) prior (uniform distribution) to reflect the initial lack of knowledge about compliance rates. The likelihood was modelled using a binomial distribution, resulting in a Beta posterior distribution. The 2022 survey data was incorporated by using the 2010 posterior as the prior for new data.

Bootstrap Analysis

The bootstrap is a resampling technique used to estimate statistics on a population by sampling a dataset with a replacement. It can be used to estimate summary statistics such as the mean or standard deviation. (Brownlee, 2024) Treatment timing compliance was assessed using bootstrap resampling techniques. From the sample of 25 treatment times, 10,000 bootstrap samples were generated to characterise uncertainty in the mean and standard deviation of treatment timing. Assuming normality, these parameters were used to estimate the probability of treatment timing falling outside the required 24–120-hour window.

Combined Risk Assessment

The overall risk of infected dogs entering the U.K. was calculated by combining the 3 probability components:

1. Infection probability during holiday $p(t)$
2. Probability of non-treatment
3. Probability of incorrect treatment timing

These were combined using the formula:

$$P(infected\ upon\ return) = P(infection)\ \times\ [P(no\ treatment) + P(wrong\ timing) - P(no\ treatment)\ \times\ P(wrong\ timing)]$$

Monte Carlo Simulation was used to propagate uncertainties through this calculation.

Introduction Risk Analysis

Annual introduction risk was evaluated by modelling the number of infected dogs a a binomial process, with parameters derived from the combined risk assessment and the number of travelling dogs. The probability of at least one introduction was calculated using:

$$P(introduction) = 1 -\ (1 - P(infected\ upon\ return))^n$$

Where:

– n – Number of dogs traveling annually

Sensitivity Analysis

Sensitivity analysis shows how different values of an independent variable affect a dependent variable under a given set of assumptions. (Kenton, 2024) The primary focus of the sensitivity analysis was on 2 critical parameters: the infection pressure (v) in France and the annual number of dogs travelling from France to Britain. These parameters were selected based on their importance to the risk assessment and the potential for their values to change.

For infection pressure, five values were tested (0.0001, 0.000187, 0.0003, 0.0004, 0.0005 weeks^-1), centred around the baseline estimate of 0.000187 weeks^-1. Similarly, five scenarios for the number of traveling dogs were evaluated (500, 1000, 2000, 3000, and 5000 dogs annually), with 1000 dogs representing the baseline scenario.

The analysis employed both linear and logarithmic transformations to investigate potential relationships between the variables. For each parameter combination, Monte Carlo Simulation with 10,000 iterations was used to generate probability distributions of annual risk of E. Multilocularis introduction. The simulation incorporated the previously calculated probabilities of infection, treatment non-compliance, and incorrect timing.

To quantify parameter sensitivity, elasticity values were calculated as the ratio of the percentage change in output to the percentage change in input parameters, relative to baseline values. The results were visualised through multiple plots including log-transformed, direct relationships and

relative change from baseline comparisons. This comprehensive approach allowed for the identification of both linear and non-linear relationships between model inputs and risk outcomes.

Chapter 3 – Results

Infection Risk during Holiday Stay

Analysis of the infection probability model revealed notable uncertainty in both the infection pressure (v) and resulting infection probability p(t). The infection pressure parameter demonstrating a mean of $1.869x10^{-4}$ under normal distribution modelling, with similar results obtained using beta distribution with a mean of $1.866x10^{-4}$. This uncertainty propagated through to the probability of infection during a 2-week holiday stay, yielding an estimated mean infection probability of $3.460x10^{-4}$ under normal distribution assumptions.

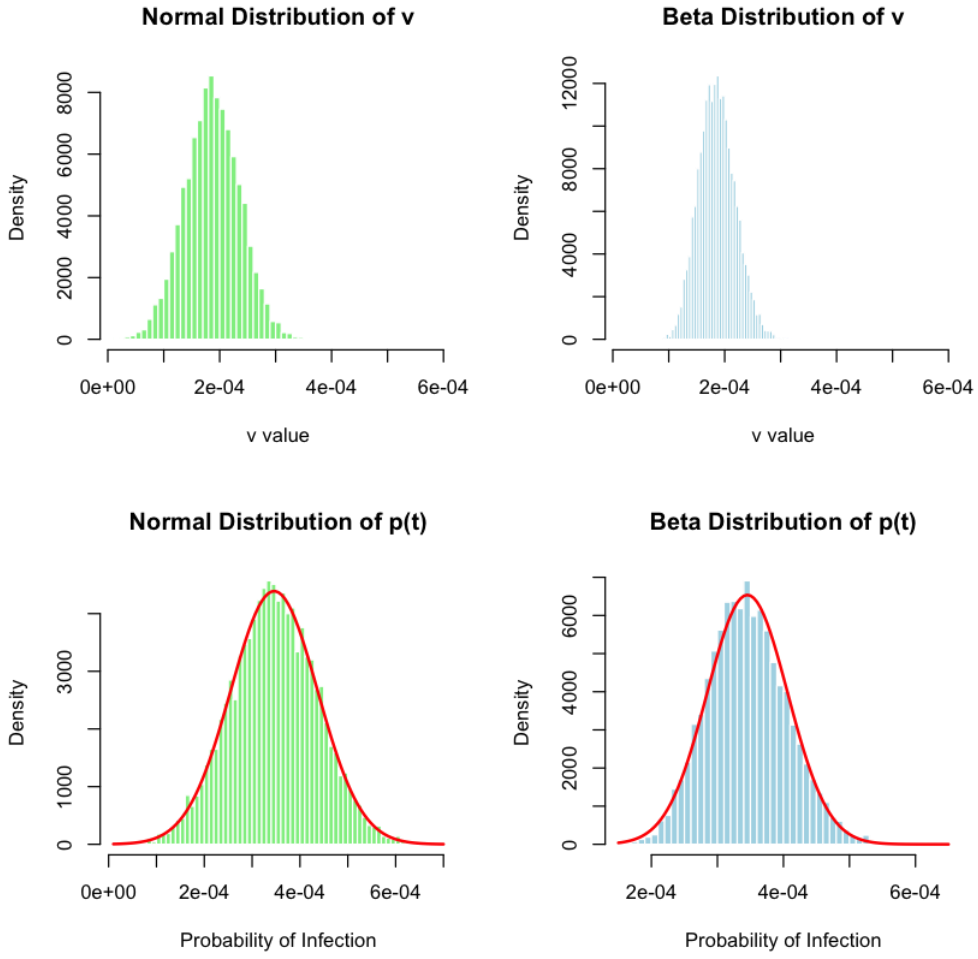


Figure 1 - Distributions of Infection Risk

Treatment Compliance Analysis

The 2010 compliance survey data, analysed using Bayesian inference with a uniform prior, indicated a non-compliance rate of approximately 10.4% with a 95% credible interval from 0.0658 to

0.1495. The posterior distribution showed a clear mode at 10%, reflecting the strong influence of the survey data on our understanding of compliance patterns. The initial analysis provided a foundation for evaluating more recent compliance trends.

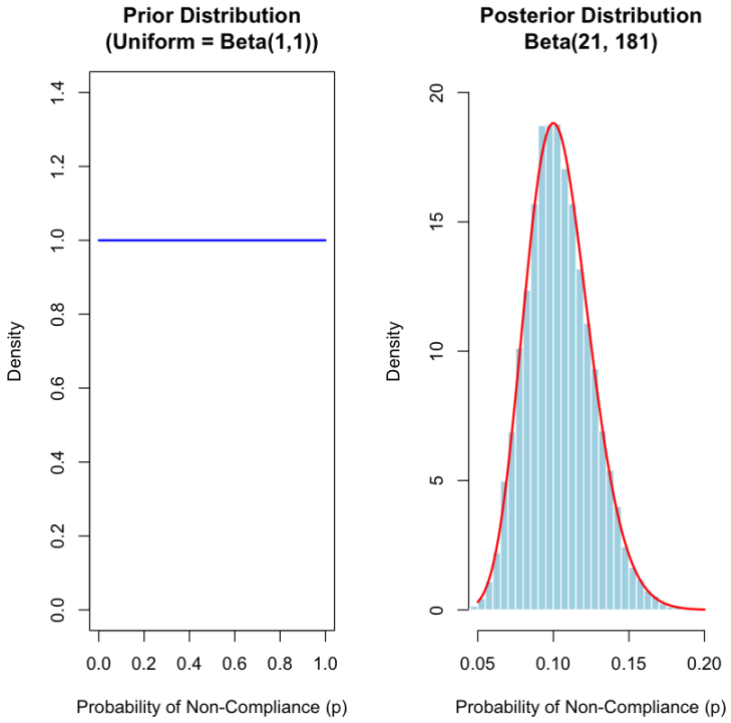


Figure 2 - Compliance Analysis Distributions

Updated Compliance Assessment

Integration of the 2022 survey data led to a refinement of compliance estimates. The updated posterior distribution indicated a slight improvement in compliance, with the mean non-compliance rate decreasing to 9.66% (95% CI: 6.12% to 13.81%). The probability of non-compliance exceeding 10% decreased to 41.32%, suggesting a potential improvement in owner adherence to treatment requirements. This update demonstrated the value of incorporating new data while retaining the information from previous assessments.

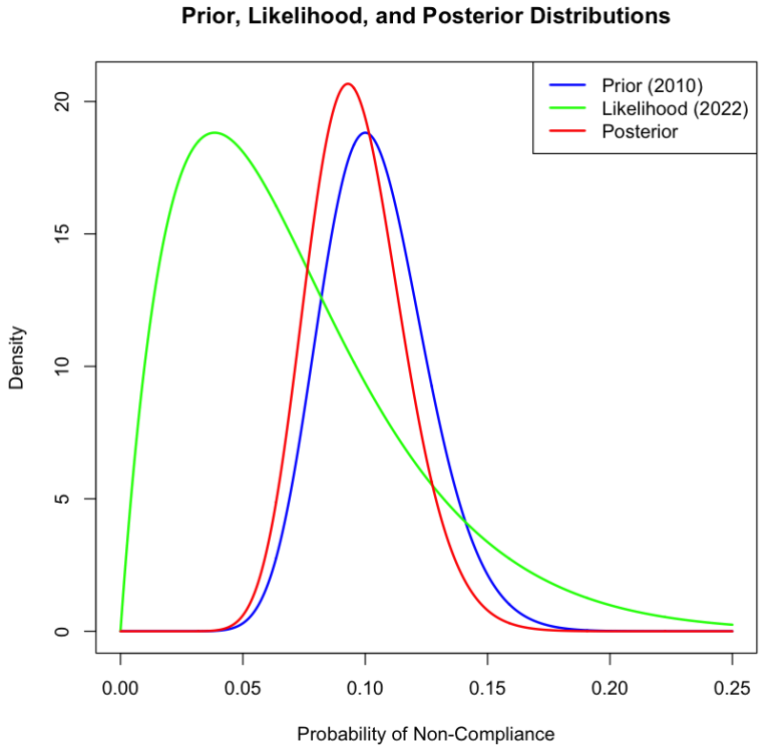


Figure 3 - Updated Compliance Distribution

Treatment Timing Patterns

Bootstrap analysis of treatment timing data revealed significant variability in when owners administered treatment relative to return travel. The mean treatment time was 71.57 hours before entry (95% CI: 57.48 to 86.16 hours), with a standard deviation of 36.53 hours (95% CI: 27.94 to 44.41 hours). Using these parameters to estimate compliance with the 24-120 hour window, approximately 19.74% (95% CI: 10.03% to 28.81%) of treatments were predicted to fall outside the required timeframe. This aligned closely with the observed sample proportion of 16% non-compliant timing.

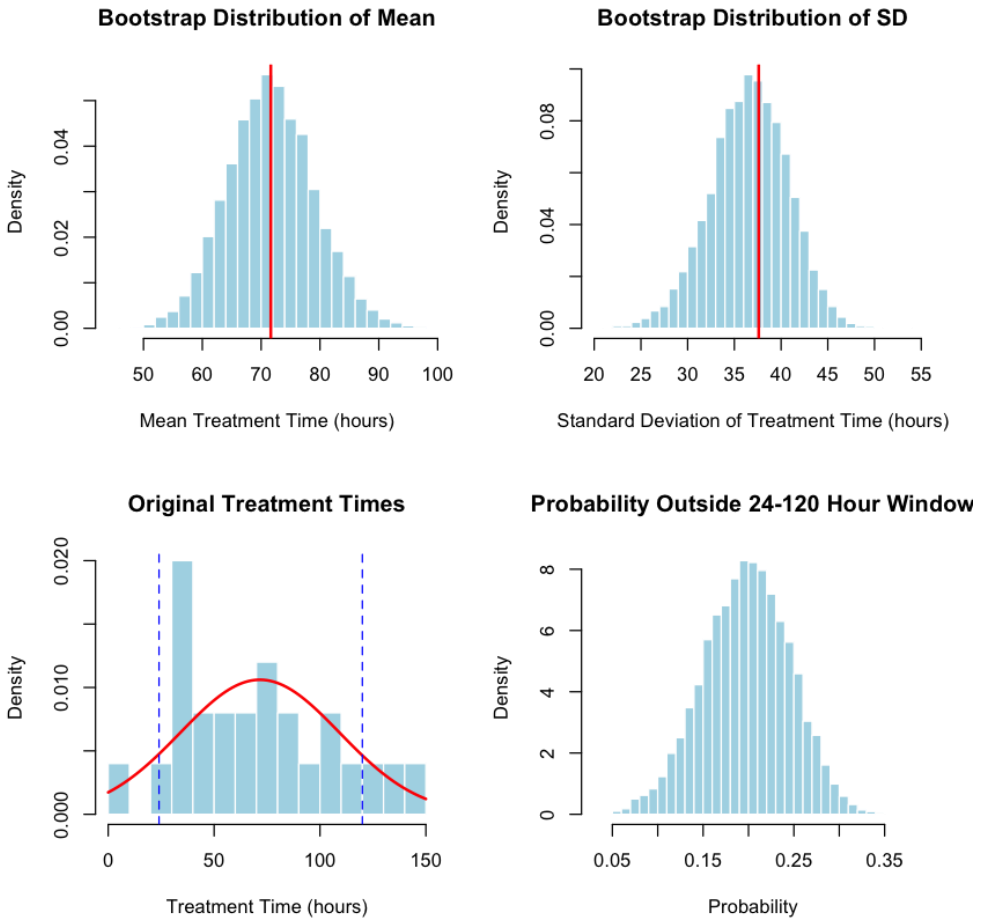


Figure 4 - Bootstrap Treatment Timing Analysis

Combined Risk Assessment

Integration of infection probability, treatment non-compliance, and timing issues yielded an overall mean probability of infection upon return of $5.153x10^{-6}$ with a 95% confidence interval of $3.506x10^{-6}$ to $6.826x10^{-6}$. Notably the analysis indicated extremely low probabilities of exceeding even modest risk thresholds, with negligible chances of surpassing 0.01%, 0.1% or even a 1% infection risk level.

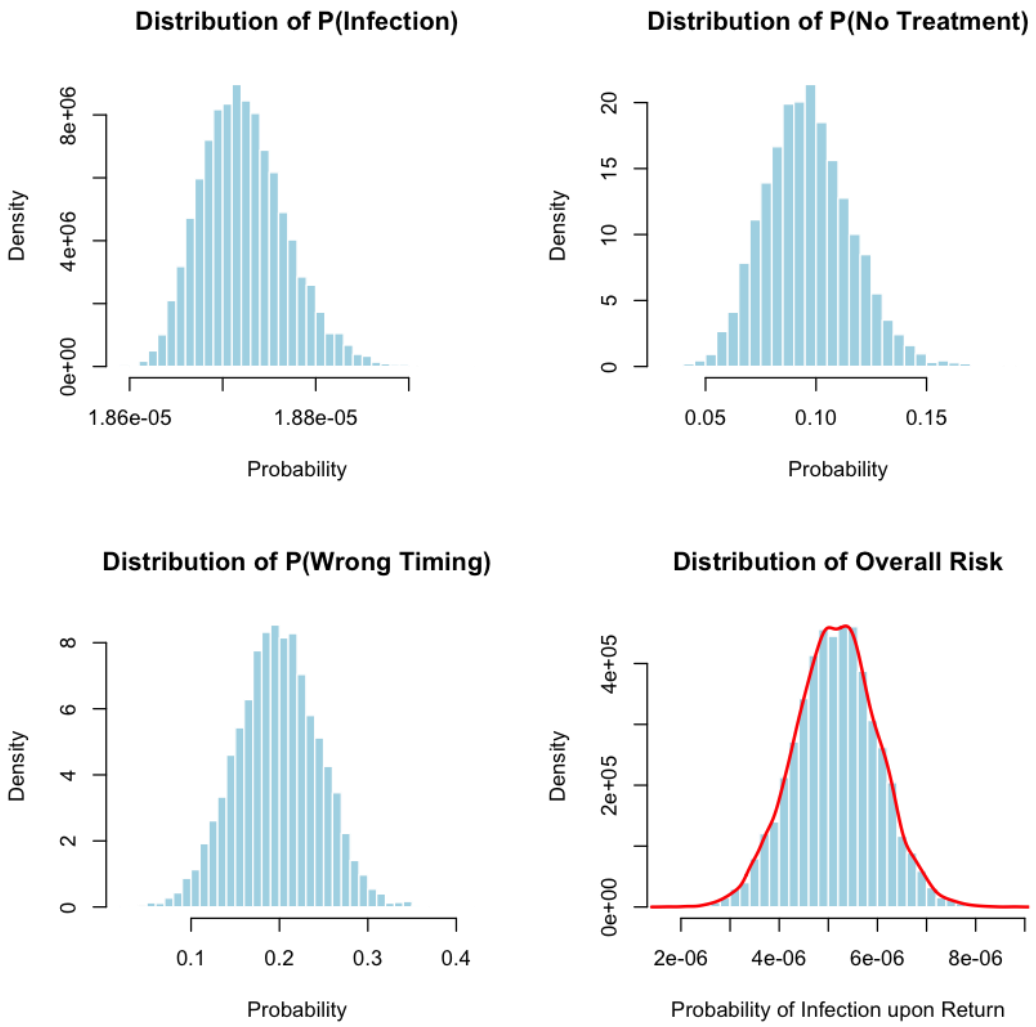


Figure 5 - Overall Infection Probability Graphs

Annual Introduction Risk

Extending the analysis to consider the annual risk of E. multilocularis introduction, based on an estimated 1,000 dogs traveling annually, produced a mean introduction probability of 0.51% (95% CI: 0.38% to 0.66%). This represents the likelihood of at least one infected dog entering Britain within a year, accounting for all compliance and timing factors.

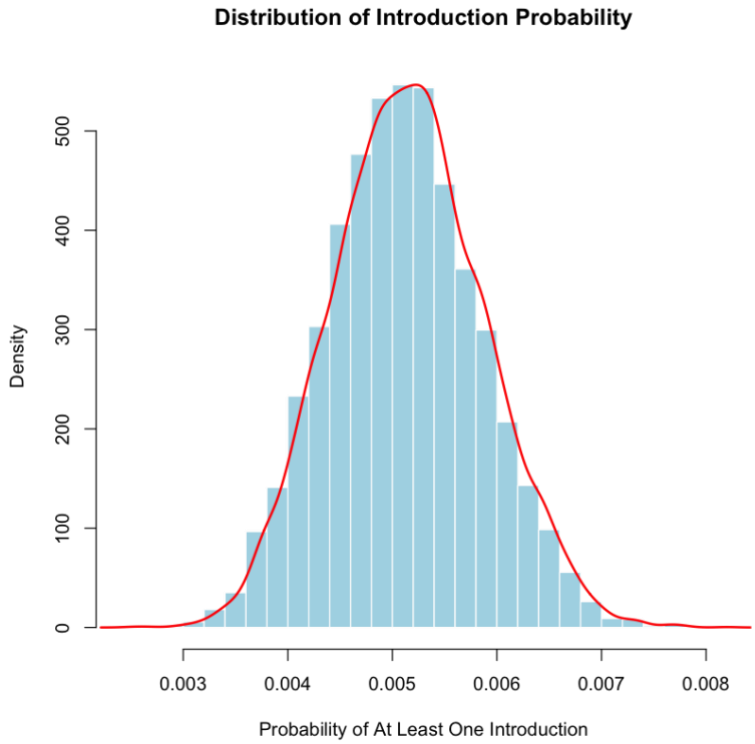


Figure 6 - Distribution at least one introduction

Parameter Sensitivity Effects

Sensitivity analysis revealed distinct relationships between key parameters and introduction risk. The infection pressure (v) demonstrated a near-linear relationship with mean annual infections when examined on a logarithmic scale, with elasticity values consistently close to 1.0 across different scenarios (ranging from 0.97 to 1.05). This indicates that changes in infection pressure produce proportional changes in introduction risk.

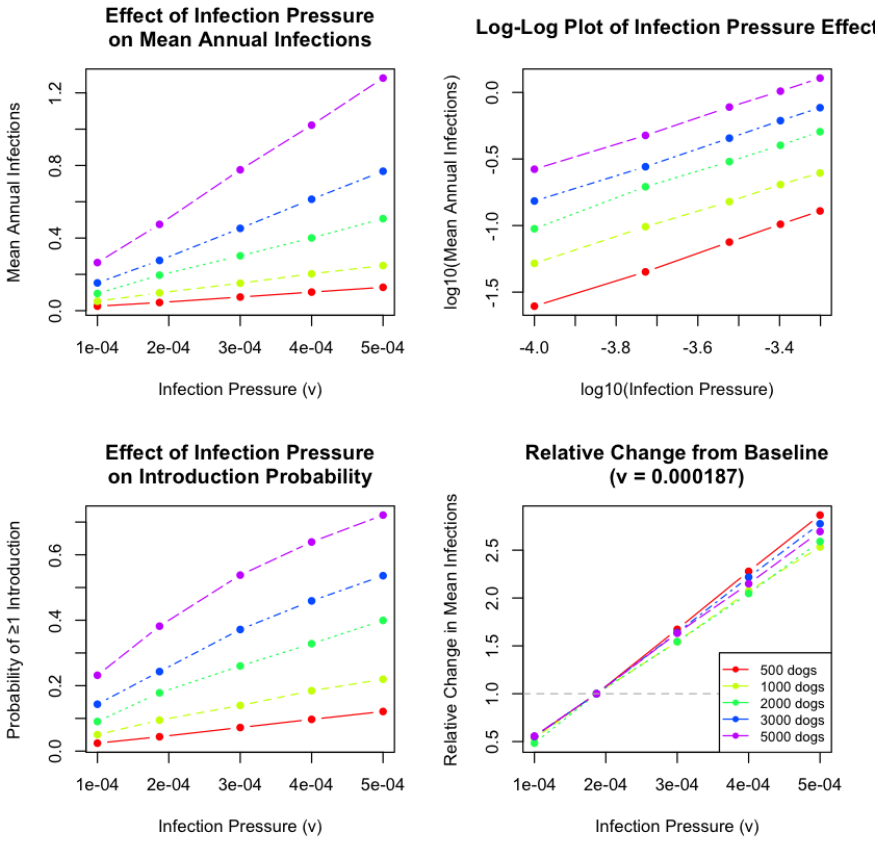


Figure 7 - Sensitivity Analysis Infection Pressure

The number of traveling dogs showed similar elasticity characteristics, with values ranging from 0.97 to 1.05 across different infection pressure scenarios. This suggests that both parameters have approximately proportional effects on introduction risk, though the absolute impact varies with baseline conditions. The analysis revealed that increasing the number of traveling dogs from 1,000 to 5,000 would increase the introduction risk roughly fivefold, while similar proportional changes were observed with variations in infection pressure.

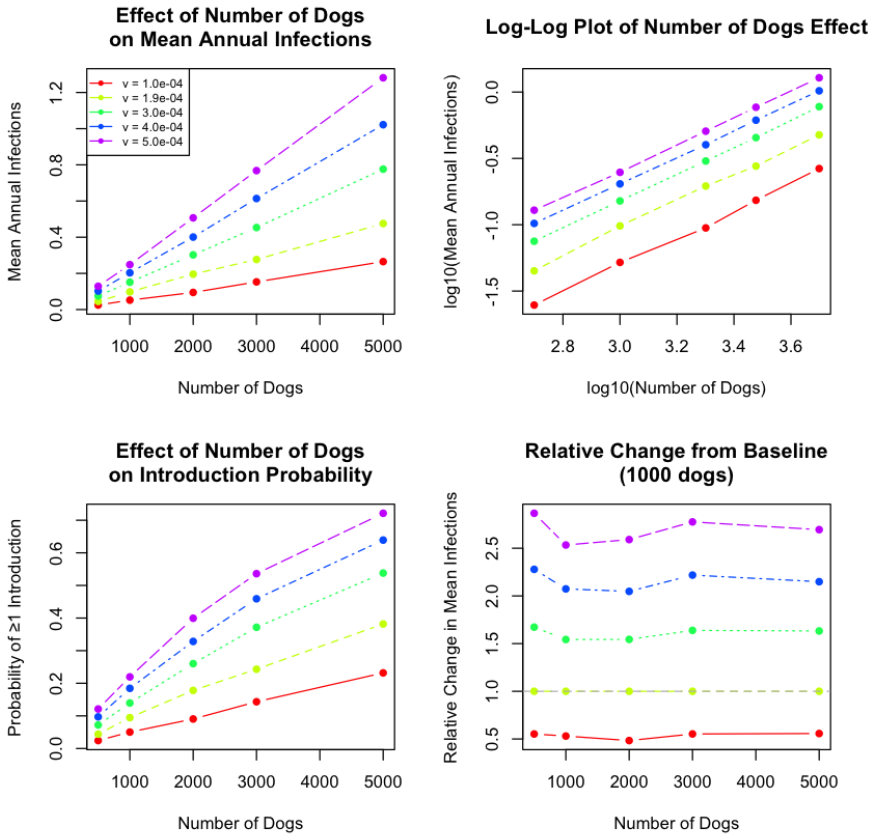


Figure 8 - Sensitivity Analysis Number of Dogs

Chapter 4 – Discussion and Conclusion

The comprehensive risk analysis of E. multilocularis introduction into Great Britain through dogs returning from France has yielded several significant findings that have important implications for public health policy and border control measures. The study revealed a relatively low overall risk of parasite introduction, with the annual probability of at least one infected dog entering Britain estimated at 0.51% (95% CI: 0.38% to 0.66%), based on current travel patterns of 1,000 dogs annually.

One of the most noteworthy findings is the relatively high level of treatment compliance among dog owners, with non-compliance rates showing a slight improvement from 10.4% in 2010 to 9.66% in 2022. This positive trend suggests that current education and enforcement measures are having some effect, though there remains room for improvement. However, the timing of treatment administration presents a more concerning picture, with approximately 19.74% of treatments falling outside the required 24-120 hour window. This higher rate of timing non-compliance compared to treatment non-compliance suggests that while owners understand the need for treatment, they may struggle with the specific timing requirements.

The sensitivity analysis provided crucial insights into how changes in key parameters might affect introduction risk. The near-linear relationships observed between both infection pressure and number of traveling dogs with introduction risk (elasticity values consistently between 0.97 and 1.05) indicate that changes in either parameter would result in proportional changes to risk. This finding has important implications for policy making, particularly as international pet travel continues to increase. The analysis suggests that doubling the number of traveling dogs would approximately double the introduction risk, highlighting the importance of maintaining effective border controls even as travel volumes increase.

Several strengths of this analysis deserve mention. The use of multiple statistical approaches - including Bayesian inference, bootstrap analysis, and Monte Carlo simulation - provided robust uncertainty quantification throughout the analysis. The integration of both historical (2010) and recent

(2022) compliance data allowed for temporal trend assessment, while the detailed treatment timing analysis offered insights into specific aspects of compliance that might need attention.

However, the study also has some limitations that should be considered. The assumption of normality in treatment timing patterns may not fully capture the true distribution of owner behavior. Additionally, the infection pressure parameter (ν) estimates rely on limited data from France, and local variations in parasite prevalence within France were not considered. The analysis also assumes that all non-compliant treatments result in the same level of risk, which may not reflect reality as the effectiveness of treatment likely varies with timing.

The findings of this study have several important public health implications. While the current risk of E. multilocularis introduction appears low, the potentially severe consequences of establishment in Britain warrant continued vigilance. The high mortality rate of untreated human cases (over 90% within 10-15 years) and the costly treatment requirements make prevention particularly important. The improvement in treatment compliance rates suggests that current regulatory frameworks are functioning, but the high rate of timing non-compliance indicates a need for better communication about timing requirements.

When comparing these findings to existing literature, our estimated non-compliance rates of 9.66% align interestingly with previous studies examining illegal entry risks. (Fisher, et al., 2023) study of seized dogs found that 27.7% of non-compliant dogs harboured endoparasites, highlighting the importance of compliance monitoring. Their finding that one non-compliant dog carried a taeniid infection, despite examining only 65 samples, supports our mathematical modelling of introduction risk. Furthermore, our timing compliance analysis revealing 19.74% of treatments falling outside the required window provides new insight into a specific aspect of compliance not previously quantified. The significant increase in traveling dogs noted in their study (from 85,786 in 2011 to 307,263 in 2019) validates our sensitivity analysis focusing on travel volumes as a key risk parameter. Their estimate of a 98% probability that one untreated dog in 10,000 would carry E. multilocularis provides an interesting comparison point to our calculated infection probability of 3.460×10^{-4}, suggesting our model's predictions are reasonably aligned with previous empirical observations.

.

References

Brownlee, J. (2024, November 10). *A Gentle Introduction to the Bootstrap Method* . Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/

Brunetti, E., & Kern, P. (2010). *Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans.* Geneva: Writing Panel for tje WHO-IWGE.

Fisher, M. A., Rees, B., Capner, C., Pritchard, S., Holdsworth, P. A., & Fitzgerald, R. A. (2023). *A survey of gastrointestinal parasites in dogs illegally entering the UK (2015–2017).* London: Veterinary Record Open.

Karpecki, P. (2024, October 26). *Myopia: A Growing Concern for Young Patients*. Retrieved from Rendia: https://rendia.com/resources/insights/myopia-growing-concern-young-patients/

Kenton, W. (2024, November 10). *What is Sensitivity Analysis*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/sensitivityanalysis.asp

McConnaughey, M. (2014). *Life Cycle of Parasites.* Greenville: East Carolina University.

ScienceDirect. (2024, October 26). *Repeasted Measures Design*. Retrieved from ScienceDirect: https://www.sciencedirect.com/topics/social-sciences/repeated-measures-design

The Editors of Encyclopaedia Britannica. (2024, November 10). *Bayesian analysis*. Retrieved from Britannica: https://www.britannica.com/science/Bayesian-analysis

Appendix

Appendix

MM925: Quantitative Risk Analysis

Packages Used

Using R version 4.40, Visual Studio Code version 1.90 and IRkernal version 1.3.2. A number of packages within R have been used listed below:

In [1]:
```r
# Set Seed - this assignment does not require student number
set.seed(123)
# Set the plot background color to white
par(bg = "white")
# Turning off errors for readability
options(warn=0)
```

Helper Functions

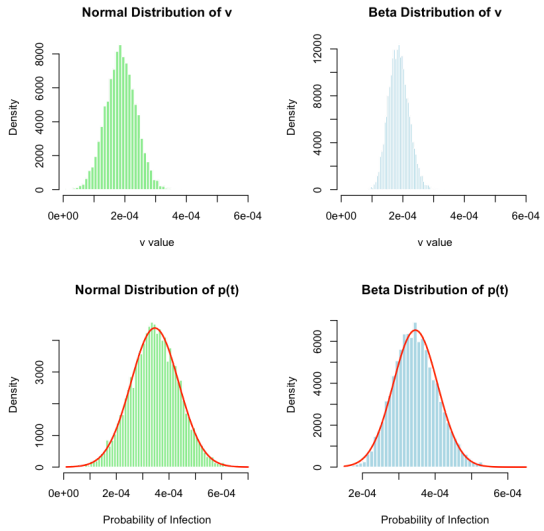
In [2]:
```r
# Core calculation functions
calculate_p <- function(v, mu, t) {
    (v * (1 - exp(-(mu + v) * t))) / (v + mu)
}

# Summary statistics helper
summarize_distribution <- function(values, name) {
    cat(sprintf("\nSummary Statistics for %s:\n", name))
    cat("Mean:", formatC(mean(values), format="e", digits=3), "\n")
    cat("Median:", formatC(median(values), format="e", digits=3), "\n")
    cat("SD:", formatC(sd(values), format="e", digits=3), "\n")
    cat("95% CI: [",
        formatC(quantile(values, 0.025), format="e", digits=3), ", ",
        formatC(quantile(values, 0.975), format="e", digits=3), "]\n")
}

# Bayesian analysis helpers
calculate_posterior_params <- function(n, s, prior_alpha = 1, prior_beta = 1) {
    list(
        alpha = s + prior_alpha,
        beta = n - s + prior_beta
    )
}

calculate_beta_mode <- function(alpha, beta) {
    if (alpha > 1 && beta > 1) {
        (alpha - 1)/(alpha + beta - 2)
    } else {
        NA  # Mode doesn't exist for some parameter combinations
    }
}
```

Question 1

In [3]:
```r
# Question 1 Data
v_min <- 0.00001
v_max <- 0.0006
v_mode <- 0.000187
mu <- 0.078
t <- 2  # 2 week holiday
n_samples <- 10000

# Normal Distribution Sample Generation
generate_normal_samples <- function(n_samples, v_min, v_max, v_mode) {
    v_sd <- (v_max - v_min) / 6  # Range rule
    v_norm <- rnorm(n_samples, mean = v_mode, sd = v_sd/2)
    # Truncate to bounds - fix the pmin/pmax syntax
    v_norm <- pmin(pmax(v_norm, v_min), v_max)  # This was incorrect before
    return(v_norm)
}

# Beta Distribution Sample Generation
generate_beta_samples <- function(n_samples, v_min, v_max, v_mode, alpha = 20) {
    mode_scaled <- (v_mode - v_min) / (v_max - v_min)
    beta <- alpha * (1 - mode_scaled) / mode_scaled

    v_beta <- rbeta(n_samples, alpha, beta)
    # Scale beta samples to the correct range
    v_beta <- v_min + (v_max - v_min) * v_beta
    return(v_beta)
}

v_norm <- generate_normal_samples(n_samples, v_min, v_max, v_mode)
v_beta <- generate_beta_samples(n_samples, v_min, v_max, v_mode)

# Calculate probabilities
p_norm <- calculate_p(v_norm, mu, t)
p_beta <- calculate_p(v_beta, mu, t)
```

```
# Create plots
par(mfrow=c(2,2))

# Plot 1: Normal Distribution of ν
hist(v_norm, breaks=50, probability=TRUE,
     main="Normal Distribution of ν",
     xlab="ν value",
     ylab="Density",
     col="lightgreen",
     border="white",
     xlim=c(v_min, v_max))

# Plot 2: Beta Distribution of ν
hist(v_beta, breaks=50, probability=TRUE,
     main="Beta Distribution of ν",
     xlab="ν value",
     ylab="Density",
     col="lightblue",
     border="white",
     xlim=c(v_min, v_max))

# Plot 3: Normal Distribution of p(t)
hist(p_norm, breaks=50, probability=TRUE,
     main="Normal Distribution of p(t)",
     xlab="Probability of Infection",
     ylab="Density",
     col="lightgreen",
     border="white")
curve(dnorm(x, mean=mean(p_norm), sd=sd(p_norm)),
      add=TRUE, col="red", lwd=2)

# Plot 4: Beta Distribution of p(t)
hist(p_beta, breaks=50, probability=TRUE,
     main="Beta Distribution of p(t)",
     xlab="Probability of Infection",
     ylab="Density",
     col="lightblue",
     border="white")
curve(dnorm(x, mean=mean(p_beta), sd=sd(p_beta)),
      add=TRUE, col="red", lwd=2)

par(mfrow=c(1,1)) # Reset plot layout

# Print summaries
cat("\n=== Normal Distribution Results ===")
summarize_distribution(v_norm, "ν (Normal)")
summarize_distribution(p_norm, "p(t) (Normal)")

cat("\n=== Beta Distribution Results ===")
summarize_distribution(v_beta, "ν (Beta)")
summarize_distribution(p_beta, "p(t) (Beta)")
```

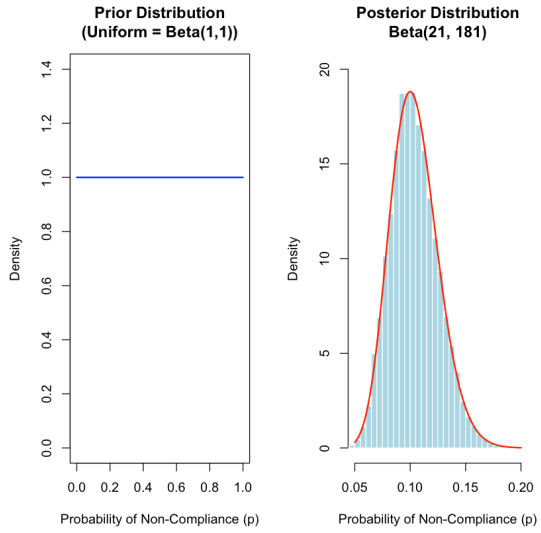
```
=== Normal Distribution Results ===
Summary Statistics for ν (Normal):
Mean: 1.869e-04
Median: 1.865e-04
SD: 4.910e-05
95% CI: [ 8.978e-05 ,  2.829e-04 ]

Summary Statistics for p(t) (Normal):
Mean: 3.460e-04
Median: 3.452e-04
SD: 9.088e-05
95% CI: [ 1.662e-04 ,  5.238e-04 ]

=== Beta Distribution Results ===
Summary Statistics for ν (Beta):
Mean: 1.866e-04
Median: 1.854e-04
SD: 3.297e-05
95% CI: [ 1.263e-04 ,  2.541e-04 ]

Summary Statistics for p(t) (Beta):
Mean: 3.454e-04
Median: 3.432e-04
SD: 6.103e-05
95% CI: [ 2.338e-04 ,  4.704e-04 ]
```



Question 2

In [4]:
```
# Question 2 Data
n <- 200    # number of trials (dogs checked)
s <- 20     # number of non-compliant cases
n_samples <- 10000

# Calculate posterior parameters
posterior_params <- calculate_posterior_params(n, s)
posterior_samples <- rbeta(n_samples, posterior_params$alpha, posterior_params$beta)

# Create plots
par(mfrow=c(1,2))

# Plot 1: Prior (Uniform = Beta(1,1))
curve(dbeta(x, 1, 1), from=0, to=1,
      main="Prior Distribution\n(Uniform = Beta(1,1))",
      xlab="Probability of Non-Compliance (p)",
      ylab="Density",
      col="blue",
      lwd=2,
      ylim=c(0, 1.4))

# Plot 2: Posterior
hist(posterior_samples, freq=FALSE, breaks=30,
     main=sprintf("Posterior Distribution\nBeta(%d, %d)",
                  posterior_params$alpha, posterior_params$beta),
     xlab="Probability of Non-Compliance (p)",
     ylab="Density",
     col="lightblue",
     border="white",
     xlim=c(0.05, 0.20),
     ylim=c(0, 20))

curve(dbeta(x, posterior_params$alpha, posterior_params$beta),
      add=TRUE, col="red", lwd=2,
      from=0.05, to=0.20)

par(mfrow=c(1,1))

# Calculate summary statistics
mode <- calculate_beta_mode(posterior_params$alpha, posterior_params$beta)
ci <- qbeta(c(0.025, 0.975), posterior_params$alpha, posterior_params$beta)

# Print summary statistics
cat(sprintf("\nPosterior Distribution (Beta(%d, %d)) Summary:\n",
           posterior_params$alpha, posterior_params$beta))
cat("Mean:", round(mean(posterior_samples), 4), "\n")
cat("Median:", round(median(posterior_samples), 4), "\n")
if(!is.na(mode)) cat("Mode:", round(mode, 4), "\n")
cat("95% Credible Interval:", round(ci[1], 4), "to", round(ci[2], 4), "\n")
```

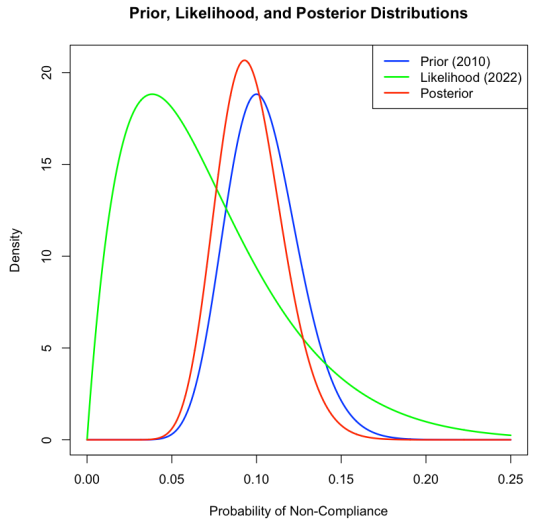
```
Posterior Distribution (Beta(21, 181)) Summary:
Mean: 0.1039
Median: 0.1022
Mode: 0.1
95% Credible Interval: 0.0658 to 0.1495
```



Question 3

In [5]:
```
# Question 3 Data
n_samples <- 10000
survey_2010 <- list(
    n = 200,  # number of trials
    s = 20    # number of non-compliant
)
survey_2022 <- list(
    n = 26,   # new trials
    s = 1     # new non-compliant
)
plot_range <- c(0, 0.25)
n_points <- 1000

# Calculate posterior parameters
prior_params <- calculate_posterior_params(survey_2010$n, survey_2010$s)
posterior_params <- calculate_posterior_params(
    n = survey_2010$n + survey_2022$n,
    s = survey_2010$s + survey_2022$s,
    prior_alpha = 1,
    prior_beta = 1
)

# Generate samples and sequence for plotting
posterior_samples <- rbeta(n_samples, posterior_params$alpha, posterior_params$beta)
x <- seq(plot_range[1], plot_range[2], length=n_points)

# Calculate densities
prior_density <- dbeta(x, prior_params$alpha, prior_params$beta)
posterior_density <- dbeta(x, posterior_params$alpha, posterior_params$beta)

# Calculate and scale likelihood
likelihood <- dbinom(survey_2022$s, size=survey_2022$n, prob=x)
likelihood_scaled <- likelihood * max(prior_density) / max(likelihood)

# Create plot
plot(x, prior_density, type="l", col="blue", lwd=2,
     main="Prior, Likelihood, and Posterior Distributions",
     xlab="Probability of Non-Compliance",
     ylab="Density",
     ylim=c(0, max(c(prior_density, posterior_density))),
     xlim=plot_range)

lines(x, likelihood_scaled, col="green", lwd=2)
lines(x, posterior_density, col="red", lwd=2)

legend("topright",
       legend=c("Prior (2010)", "Likelihood (2022)", "Posterior"),
       col=c("blue", "green", "red"),
       lwd=2)

# Print summaries using shared summarize_distribution function
cat("\nPosterior Distribution Summary Statistics:")
summarize_distribution(posterior_samples, "Posterior")

# Calculate comparison statistics
mode <- calculate_beta_mode(posterior_params$alpha, posterior_params$beta)
ci <- qbeta(c(0.025, 0.975), posterior_params$alpha, posterior_params$beta)
prior_mean <- prior_params$alpha/(prior_params$alpha + prior_params$beta)
posterior_mean <- posterior_params$alpha/(posterior_params$alpha + posterior_params$beta)
prob_greater_10 <- mean(posterior_samples > 0.10)

cat("\nComparison of Estimates:\n")
cat("Prior Mean (from 2010):", round(prior_mean, 4), "\n")
cat("Posterior Mean (after 2022):", round(posterior_mean, 4), "\n")
cat("Change in estimate:", round(posterior_mean - prior_mean, 4), "\n")
cat("\nProbability that non-compliance rate > 10%:", round(prob_greater_10, 4), "\n")
```

```
Posterior Distribution Summary Statistics:
Summary Statistics for Posterior:
Mean: 9.663e-02
Median: 9.566e-02
SD: 1.966e-02
95% CI: [ 6.120e-02 ,  1.381e-01 ]

Comparison of Estimates:
Prior Mean (from 2010): 0.104
Posterior Mean (after 2022): 0.0965
Change in estimate: -0.0075

Probability that non-compliance rate > 10%: 0.4132
```



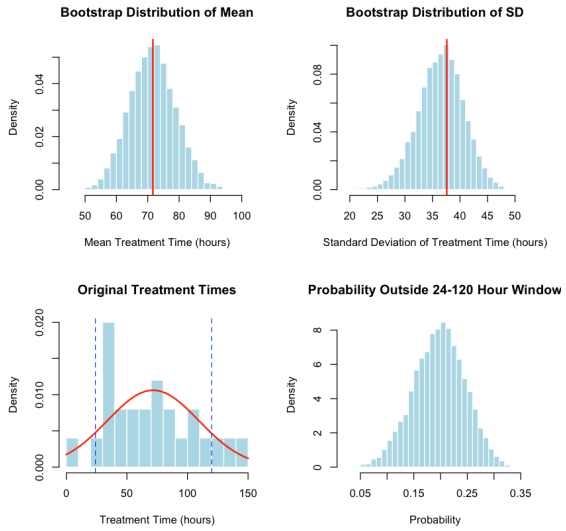
Question 4

In [10]:

```r
# Data for Question 4
treatment_times <- c(105, 73, 140, 53, 100, 39, 129, 41, 115, 55, 145, 89, 45,
                     35, 7, 36, 29, 109, 70, 31, 76, 90, 64, 39, 76)

# Compliance window
lower_limit <- 24
upper_limit <- 120

# Bootstrap function
bootstrap_samples <- function(data, n_boot=10000) {
    n <- length(data)
    means <- numeric(n_boot)
    sds <- numeric(n_boot)
    probs_outside <- numeric(n_boot)

    for(i in 1:n_boot) {
        boot_sample <- sample(data, size=n, replace=TRUE)
        means[i] <- mean(boot_sample)
        sds[i] <- sd(boot_sample)
        prob_below <- pnorm(lower_limit, means[i], sds[i])
        prob_above <- 1 - pnorm(upper_limit, means[i], sds[i])
        probs_outside[i] <- prob_below + prob_above
    }

    return(list(means=means, sds=sds, probs_outside=probs_outside))
}

# Run bootstrap
boot_results <- bootstrap_samples(treatment_times)

# Create visualizations
par(mfrow=c(2,2))

# Plot 1: Bootstrap distribution of means
hist(boot_results$means, breaks=30, probability=TRUE,
     main="Bootstrap Distribution of Mean",
     xlab="Mean Treatment Time (hours)",
     ylab="Density",
     col="lightblue", border="white")
abline(v=mean(treatment_times), col="red", lwd=2)

# Plot 2: Bootstrap distribution of standard deviations
hist(boot_results$sds, breaks=30, probability=TRUE,
     main="Bootstrap Distribution of SD",
     xlab="Standard Deviation of Treatment Time (hours)",
     ylab="Density",
     col="lightblue", border="white")
abline(v=sd(treatment_times), col="red", lwd=2)

# Plot 3: Original data with normal fit
hist(treatment_times, breaks=10, probability=TRUE,
     main="Original Treatment Times",
     xlab="Treatment Time (hours)",
     ylab="Density",
```

```
     col="lightblue", border="white")
curve(dnorm(x, mean(treatment_times), sd(treatment_times)),
      add=TRUE, col="red", lwd=2)
abline(v=c(lower_limit, upper_limit), col="blue", lty=2)

# Plot 4: Bootstrap distribution of probability outside window
hist(boot_results$probs_outside, breaks=30, probability=TRUE,
     main="Probability Outside 24-120 Hour Window",
     xlab="Probability",
     ylab="Density",
     col="lightblue", border="white")

par(mfrow=c(1,1))

# Print summaries using shared summarize_distribution function
summarize_distribution(boot_results$means, "Bootstrap Mean Treatment Time")
summarize_distribution(boot_results$sds, "Bootstrap Standard Deviation")
summarize_distribution(boot_results$probs_outside, "Probability Outside Window")

# Additional statistics
actual_outside <- mean(treatment_times < lower_limit | treatment_times > upper_limit)
cat("\nActual proportion outside window in sample:", round(actual_outside, 4), "\n")
```

```
Summary Statistics for Bootstrap Mean Treatment Time:
Mean: 7.157e+01
Median: 7.152e+01
SD: 7.366e+00
95% CI: [ 5.748e+01 ,  8.616e+01 ]

Summary Statistics for Bootstrap Standard Deviation:
Mean: 3.653e+01
Median: 3.672e+01
SD: 4.175e+00
95% CI: [ 2.794e+01 ,  4.441e+01 ]

Summary Statistics for Probability Outside Window:
Mean: 1.974e-01
Median: 1.989e-01
SD: 4.806e-02
95% CI: [ 1.003e-01 ,  2.881e-01 ]

Actual proportion outside window in sample: 0.16
```



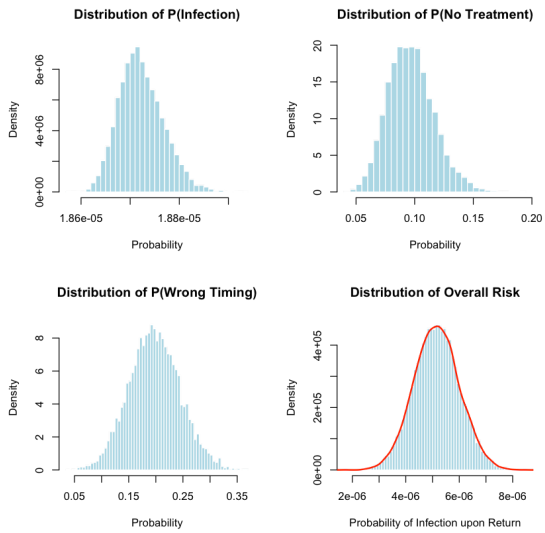
Question 5

In [26]:

```
# Data for Question 5
n_samples <- 10000
infection_params <- list(
    alpha = 20,
    v_min = 0.00001,
    v_max = 0.0006,
    v_mode = 0.000187,
    mu = 0.078,
    t = 2
)
treatment_params <- list(
    alpha = 22,
    beta = 206
)
timing_params <- list(
    mean = 0.1982,
    sd = 0.0472
)

# Generate infection probability samples
beta_q1 <- infection_params$alpha * (1 - infection_params$v_mode)/infection_params$v_mode
p1_samples <- rbeta(n_samples, infection_params$alpha, beta_q1)
p1_samples <- infection_params$v_min + (infection_params$v_max - infection_params$v_min) * p1_samples
p1_samples <- calculate_p(p1_samples, infection_params$mu, infection_params$t)
```

```r
# Generate treatment probability samples
p2_samples <- rbeta(n_samples, treatment_params$alpha, treatment_params$beta)

# Generate timing probability samples
p3_samples <- rnorm(n_samples, timing_params$mean, timing_params$sd)
p3_samples <- pmin(pmax(p3_samples, 0), 1)

# Calculate overall probability
overall_prob <- p1_samples * (p2_samples + p3_samples - p2_samples * p3_samples)

# Create visualization
par(mfrow=c(2,2))

# Plot individual components
hist(p1_samples, breaks=50, probability=TRUE,
     main="Distribution of P(Infection)",
     xlab="Probability",
     ylab="Density",
     col="lightblue", border="white")

hist(p2_samples, breaks=50, probability=TRUE,
     main="Distribution of P(No Treatment)",
     xlab="Probability",
     ylab="Density",
     col="lightblue", border="white")

hist(p3_samples, breaks=50, probability=TRUE,
     main="Distribution of P(Wrong Timing)",
     xlab="Probability",
     ylab="Density",
     col="lightblue", border="white")

hist(overall_prob, breaks=50, probability=TRUE,
     main="Distribution of Overall Risk",
     xlab="Probability of Infection upon Return",
     ylab="Density",
     col="lightblue", border="white")
lines(density(overall_prob), col="red", lwd=2)

par(mfrow=c(1,1))

# Print summaries using shared summarize_distribution function
summarize_distribution(overall_prob, "Overall Infection Probability")

# Additional probability thresholds
thresholds <- c(0.0001, 0.001, 0.01)
cat("\nProbability that infection risk exceeds:\n")
for(threshold in thresholds) {
    pct <- threshold * 100
    prob <- mean(overall_prob > threshold)
    cat(sprintf("%.3f%%: %.4f\n", pct, prob))
}
```

```
Summary Statistics for Overall Infection Probability:
Mean: 5.153e-06
Median: 5.150e-06
SD: 8.480e-07
95% CI: [ 3.506e-06 ,  6.826e-06 ]

Probability that infection risk exceeds:
0.010%: 0.0000
0.100%: 0.0000
1.000%: 0.0000
```



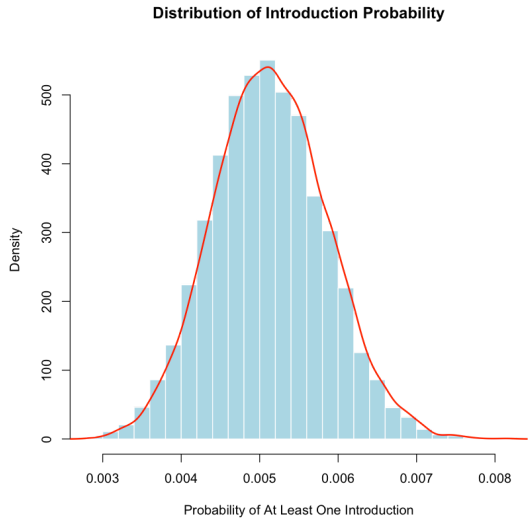
Question 6

In [60]:

```r
# Multiple simulations to get distribution of introduction probability
n_dogs <- 1000
n_sims <- 10000  # Number of simulations for each probability calculation
n_iterations <- 10000  # Number of probability estimates to generate distribution
```

```
# Calculate probability of introduction multiple times
introduction_probs <- replicate(n_iterations, {
    # For each iteration, simulate 1000 dogs and calculate probability of at least one introduction
    yearly_sim <- rbinom(n_sims, size=n_dogs, prob=overall_prob)
    mean(yearly_sim > 0)  # Probability of at least one introduction
})

# Create visualization
hist(introduction_probs,
     breaks=30,
     probability=TRUE,
     main="Distribution of Introduction Probability",
     xlab="Probability of At Least One Introduction",
     ylab="Density",
     col="lightblue",
     border="white")

# Add density curve
lines(density(introduction_probs), col="red", lwd=2)

# Print summary statistics
cat("\nDistribution of E. multilocularis Introduction Probability:\n")
cat("Mean probability:", round(mean(introduction_probs), 4), "\n")
cat("Median probability:", round(median(introduction_probs), 4), "\n")
cat("95% CI: [",
    round(quantile(introduction_probs, 0.025), 4), ", ",
    round(quantile(introduction_probs, 0.975), 4), "]\n")
```

```
Distribution of E. multilocularis Introduction Probability:
Mean probability: 0.0051
Median probability: 0.0051
95% CI: [ 0.0038 ,  0.0066 ]
```

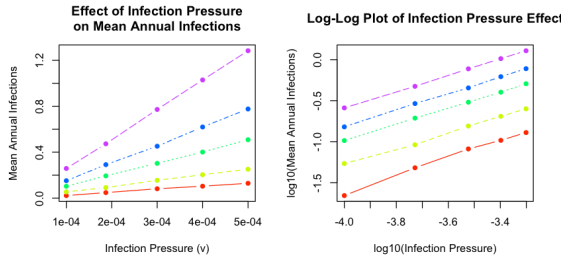


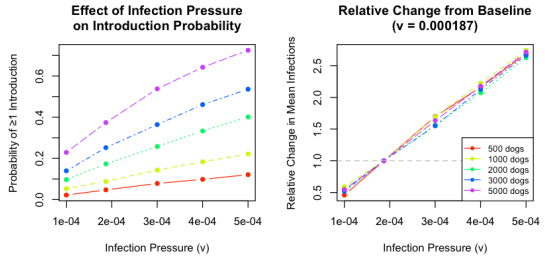
Question 7

In [74]:

```
# Create scenarios for sensitivity analysis
v_scenarios <- c(0.0001, 0.000187, 0.0003, 0.0004, 0.0005)
n_dogs_scenarios <- c(500, 1000, 2000, 3000, 5000)

# Create matrices to store results
mean_matrix <- matrix(NA, nrow=length(v_scenarios), ncol=length(n_dogs_scenarios))
prob_matrix <- matrix(NA, nrow=length(v_scenarios), ncol=length(n_dogs_scenarios))

# Calculate results
for(i in seq_along(v_scenarios)) {
    for(j in seq_along(n_dogs_scenarios)) {
        p1 <- calculate_p(v_scenarios[i], mu, t)
        p2 <- posterior_mean
        p3 <- mean(p3_samples)
        overall_prob <- p1 * (p2 + p3 - p2 * p3)
        annual_infections <- rbinom(n_samples, size=n_dogs_scenarios[j], prob=overall_prob)
        mean_matrix[i,j] <- mean(annual_infections)
        prob_matrix[i,j] <- mean(annual_infections > 0)
    }
}

# Part 1: Effect of Infection Pressure (v)
par(mfrow=c(2,2))

# 1. Linear relationship for different dog populations
matplot(v_scenarios, mean_matrix, type="b", pch=16,
        col=rainbow(length(n_dogs_scenarios)),
        main="Effect of Infection Pressure\non Mean Annual Infections",
        xlab="Infection Pressure (v)",
        ylab="Mean Annual Infections")

# 2. Log-transformed relationship
matplot(log10(v_scenarios), log10(mean_matrix), type="b", pch=16,
        col=rainbow(length(n_dogs_scenarios)),
```

```
        main="Log-Log Plot of Infection Pressure Effect",
        xlab="log10(Infection Pressure)",
        ylab="log10(Mean Annual Infections)")

# 3. Probability of at least one infection vs ν
matplot(ν_scenarios, prob_matrix, type="b", pch=16,
        col=rainbow(length(n_dogs_scenarios)),
        main="Effect of Infection Pressure\non Introduction Probability",
        xlab="Infection Pressure (ν)",
        ylab="Probability of ≥1 Introduction")

# 4. Relative change from baseline
baseline_ν <- 0.000187
relative_change <- sweep(mean_matrix, 2, mean_matrix[which(ν_scenarios == baseline_ν),], "/")
matplot(ν_scenarios, relative_change, type="b", pch=16,
        col=rainbow(length(n_dogs_scenarios)),
        main="Relative Change from Baseline\n(ν = 0.000187)",
        xlab="Infection Pressure (ν)",
        ylab="Relative Change in Mean Infections")
abline(h=1, lty=2, col="gray")
legend("bottomright", legend=paste(n_dogs_scenarios, "dogs"),
       col=rainbow(length(n_dogs_scenarios)), pch=16, lty=1,
       cex=0.8, bg="white")

# Part 2: Effect of Number of Dogs
par(mfrow=c(2,2))

# 1. Linear relationship for different infection pressures
matplot(n_dogs_scenarios, t(mean_matrix), type="b", pch=16,
        col=rainbow(length(ν_scenarios)),
        main="Effect of Number of Dogs\non Mean Annual Infections",
        xlab="Number of Dogs",
        ylab="Mean Annual Infections")
legend("topleft", legend=paste("ν =", formatC(ν_scenarios, format="e", digits=1)),
       col=rainbow(length(ν_scenarios)), pch=16, lty=1,
       cex=0.7, bg="white")

# 2. Log-transformed relationship
matplot(log10(n_dogs_scenarios), log10(t(mean_matrix)), type="b", pch=16,
        col=rainbow(length(ν_scenarios)),
        main="Log-Log Plot of Number of Dogs Effect",
        xlab="log10(Number of Dogs)",
        ylab="log10(Mean Annual Infections)")

# 3. Probability of at least one infection vs number of dogs
matplot(n_dogs_scenarios, t(prob_matrix), type="b", pch=16,
        col=rainbow(length(ν_scenarios)),
        main="Effect of Number of Dogs\non Introduction Probability",
        xlab="Number of Dogs",
        ylab="Probability of ≥1 Introduction")

# 4. Relative change from baseline
baseline_n <- 1000
relative_change <- sweep(t(mean_matrix), 1, t(mean_matrix)[,which(n_dogs_scenarios == baseline_n)], "/")
matplot(n_dogs_scenarios, relative_change, type="b", pch=16,
        col=rainbow(length(ν_scenarios)),
        main="Relative Change from Baseline\n(1000 dogs)",
        xlab="Number of Dogs",
        ylab="Relative Change in Mean Infections")
abline(h=1, lty=2, col="gray")

# Reset plot layout
par(mfrow=c(1,1))

# Print summary statistics for number of dogs effect
cat("\nEffect of Number of Dogs:\n")
# Calculate elasticity for each infection pressure
for(i in seq_along(ν_scenarios)) {
    model <- lm(log10(mean_matrix[i,]) ~ log10(n_dogs_scenarios))
    cat(sprintf("\nFor ν = %e:\n", ν_scenarios[i]))
    cat("Elasticity (log-log slope):", round(coef(model)[2], 4), "\n")
    cat("R-squared:", round(summary(model)$r.squared, 4), "\n")
}
```






```
Effect of Number of Dogs:

For ν = 1.000000e-04:
Elasticity (log-log slope): 1.0483
R-squared: 0.9954

For ν = 1.870000e-04:
Elasticity (log-log slope): 1.0052
R-squared: 0.9993

For ν = 3.000000e-04:
Elasticity (log-log slope): 0.9739
R-squared: 0.9995

For ν = 4.000000e-04:
Elasticity (log-log slope): 0.9978
R-squared: 0.9998

For ν = 5.000000e-04:
Elasticity (log-log slope): 1.0021
R-squared: 0.9998
```