

Balcaskie Monitoring 2025

Petra Guy

2026-01-20

INTERNAL DOCUMENT ONLY

SUMMARY Average change in height for oak was 44% compared to control. To test significance data was transformed using 0.25 power transformation to improve normality and homogeneity of variance. Analysis of transformed data showed change in height due to pellet not significant ($p = 0.1$, ANOVA).

OR... Change in (median) height of oak due to pellet was modeled to be 19% greater than untreated trees, (there is 95% probability that this value lies in the interval 9.5% - 74%) with an 89% probability that the effect is positive. (Moderate support for positive effect if we round to 90%!)

OR... The result above is modeled on the data with all zeros removed (for log transform) but - I would argue that these zeros hold information which is then lost. If we include the zeros in the model - by making them 0.1, we get... Change in (median) height of oak due to pellet was modeled to be 40% greater compared to untreated trees with 97% probability that the effect is positive. There is 95% probability that this effect lies within the interval 0.3% to 96%

INTRODUCTION Balcaskie is ex-ag site which was split into two areas by customer, one area was control, the other half of the field was treated. Two emf were supplied, but we dont know what went where. MBR planting, but only trees in sufficient number to be retained for analysis were beech and oak.

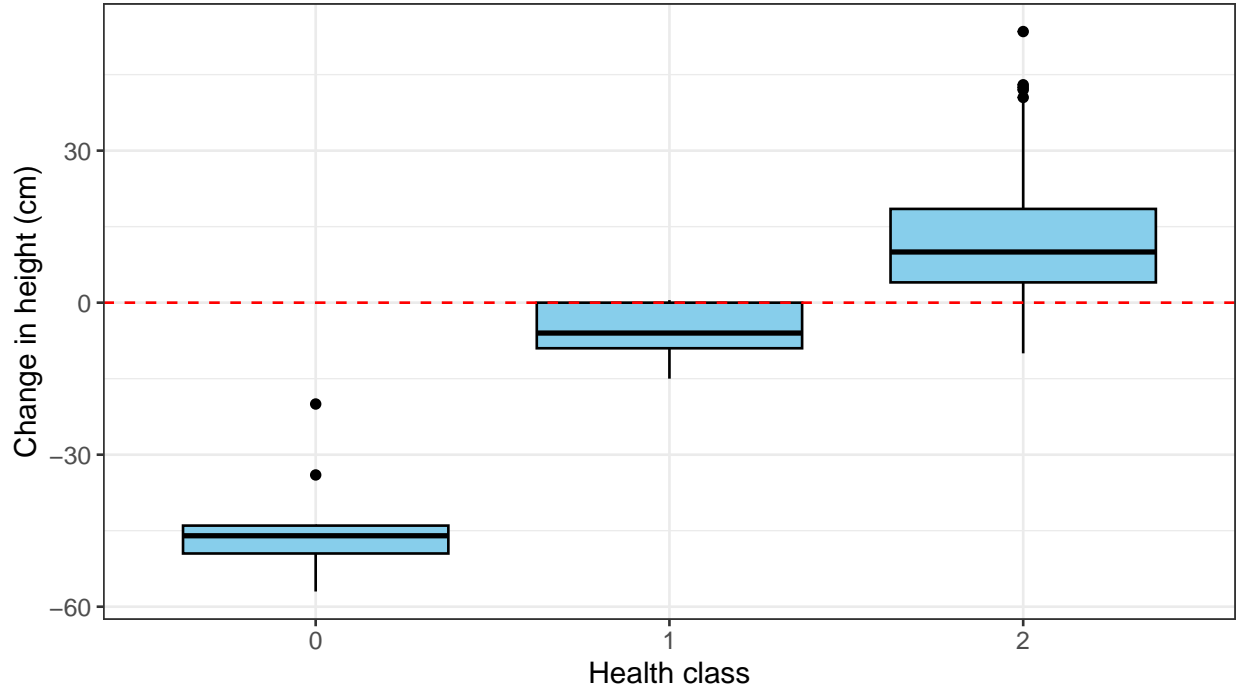
TREATMENTS Pellet and control

METHODS heights and rcd taken in autumn 2024 and again autumn 2025

ANALYSIS Because there are no repeats/randomization - so arguable we should do plot averages - but we will consider trees as independent. Data is tree-tree mean change in height for ~ 12 months. We will explore mean change in heights per plot and test for significance using ANOVA. We will also model change in height using Bayesian GLM.

Have not bothered with Fagus as exploration showed little/no effect.

ALso not done rcd as decided to move on. Can come back to that later



Data check, negative delta_h for health_0 and health_1 trees as expected - some health_2 trees as well though - these will be removed from the data for change in height.

The table below summarizes the number of trees with positive delta_h which will be used in change in height plots, delta h ≥ 0 . There are roughly equal numbers of oak and beech in the control and treatment areas after -ves removed.

Table 1: Number of trees in each treatment by health class

Tree Species	Health	n(Control)	n(Pellet)
Fagus_sylvatica	2	93	89
Fagus_sylvatica	1	0	1
Quercus_robur	1	1	0
Quercus_robur	2	93	93

Table 2: Mean change in heights (cm)

Tree Species	Treatment	mean_delta_h	se	n
Fagus_sylvatica	Control	17.956989	0.9712636	93
Fagus_sylvatica	Pellet	16.905556	1.2307416	90
Quercus_robur	Control	6.835106	0.5193790	94
Quercus_robur	Pellet	9.838710	0.9163264	93

Tables below show summary of number of trees per plot and mean values when we omit the zeros, i.e delta h > 0 - this is the data subset used for the ANOVA and GLM. Note this shifts the mean up a bit and we have lost 5 trees from control and 1 from pellet. Possibly these losses could influence the data since there are more in control than in pellet. If these were brought back in at say 0.1, then $\ln(0.1) = -2.3$, there are 5 of these in control and only 1 in pellet - hence the mean for control would be reduced more than that

of pellet. Even if we used $\ln(1) = 0$ could have an impact? Since the zero change in heights do represent a physical difference between the plots - more had 0 change in control than pellet

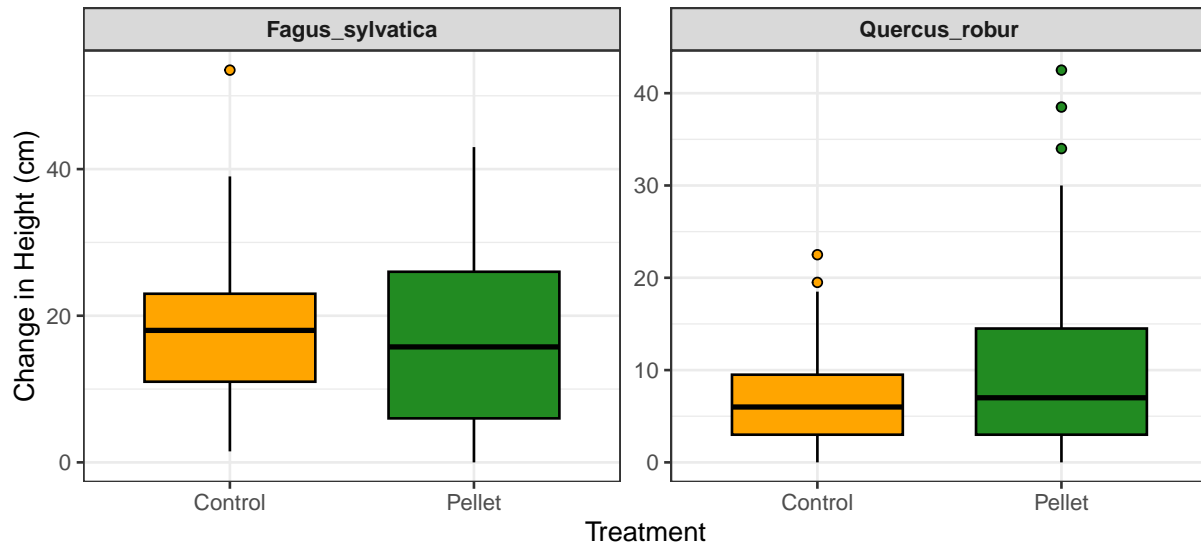
Table 3: Number of trees in each treatment by health class

Tree Species	Health	n(Control)	n(Pellet)
Fagus_sylvatica	2	93	86
Quercus_robur	1	1	0
Quercus_robur	2	88	92

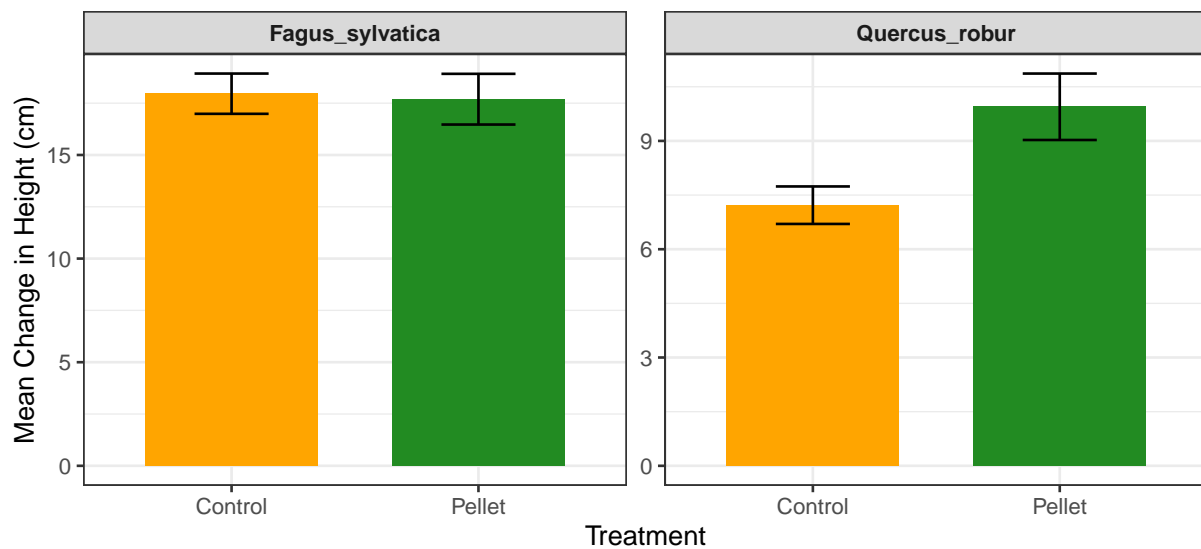
Table 4: Mean change in heights (cm)

Tree Species	Treatment	mean_delta_h	se	n
Fagus_sylvatica	Control	17.956989	0.9712636	93
Fagus_sylvatica	Pellet	17.691861	1.2231569	86
Quercus_robur	Control	7.219101	0.5192266	89
Quercus_robur	Pellet	9.945652	0.9200109	92

Distribution of delta_h by Treatment and Tree Species



Mean delta_h SE by Treatment and Tree Species



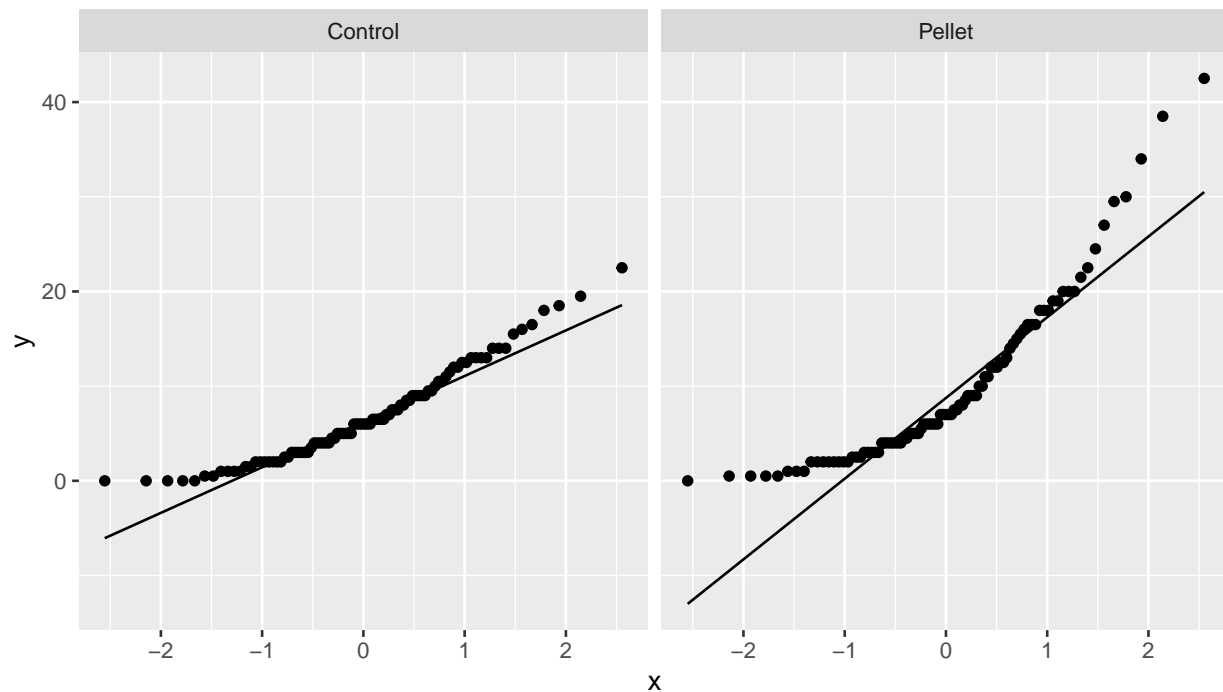
The charts show the distribution of mean change in height, and bar plot for mean value with se - we can see from se that the change for oak is possibly significant? (PS, this data includes the 0 change in heights)

First table shows shapiro wilk on un-transformed delta_h. Second shows Levine test of homogeneity of variance. Showing we need transformation

```
## # A tibble: 2 x 2
##   Treatment    shapiro_p
##   <chr>      <dbl>
## 1 Control    0.000308
## 2 Pellet    0.0000000499

##
## F test to compare two variances
```

```
##
## data: delta_h by Treatment
## F = 0.32472, num df = 93, denom df = 92, p-value = 1.386e-07
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.2154308 0.4892367
## sample estimates:
## ratio of variances
## 0.3247231
```



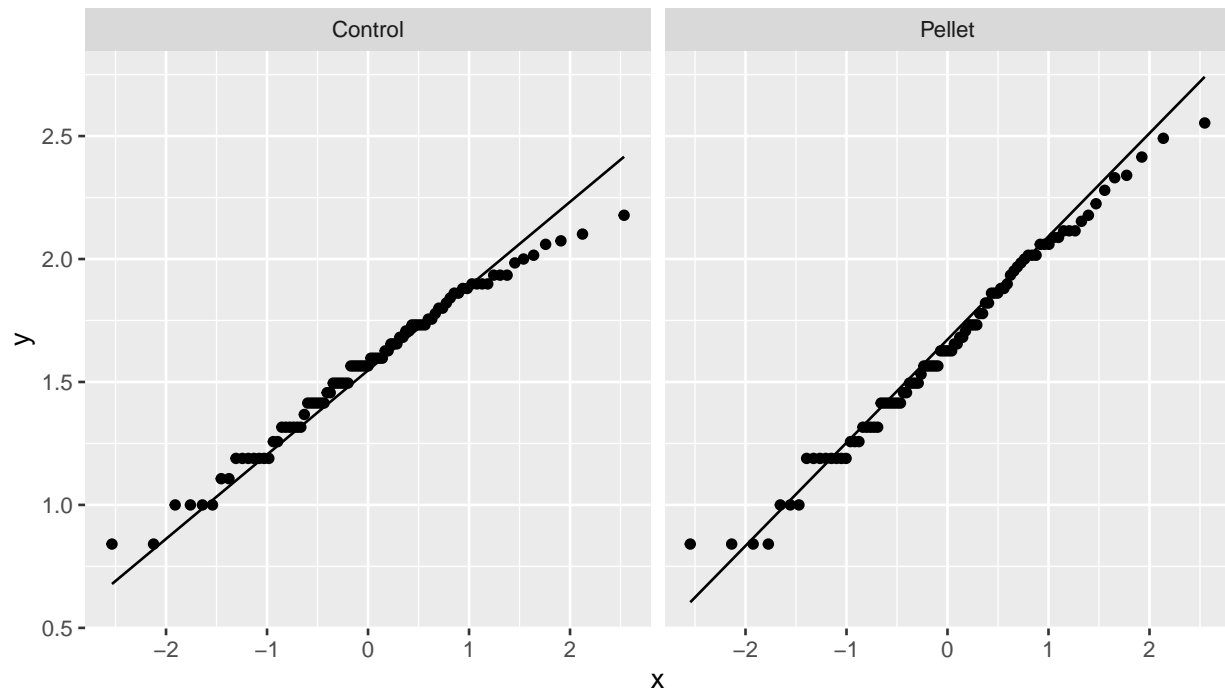
QQplots for untransformed data showing lack of normality. Data is not Normal and heteroskedastic, since its +ve only change in height - this is expected. Log transform was tried, but data still showed signs of non-normality

Box-cox suggest transform of $x^{0.25}$. The first table shows the shapiro wilk for normality on the transformed data, the second is the test for homog of variances. Plots are qqplots of transformed data. Finally - summary(anova) for $\text{delta}_h^{0.25}$.

```
## # A tibble: 2 x 2
##   Treatment shapiro_p
##   <chr>      <dbl>
## 1 Control    0.298
## 2 Pellet    0.476
```

```
##
## F test to compare two variances
##
## data: loght by Treatment
## F = 0.66042, num df = 88, denom df = 91, p-value = 0.05161
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```
## 0.4355259 1.0029232
## sample estimates:
## ratio of variances
## 0.6604205
```



```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    1  0.333   0.3326   2.646   0.106
## Residuals  179 22.497   0.1257
```

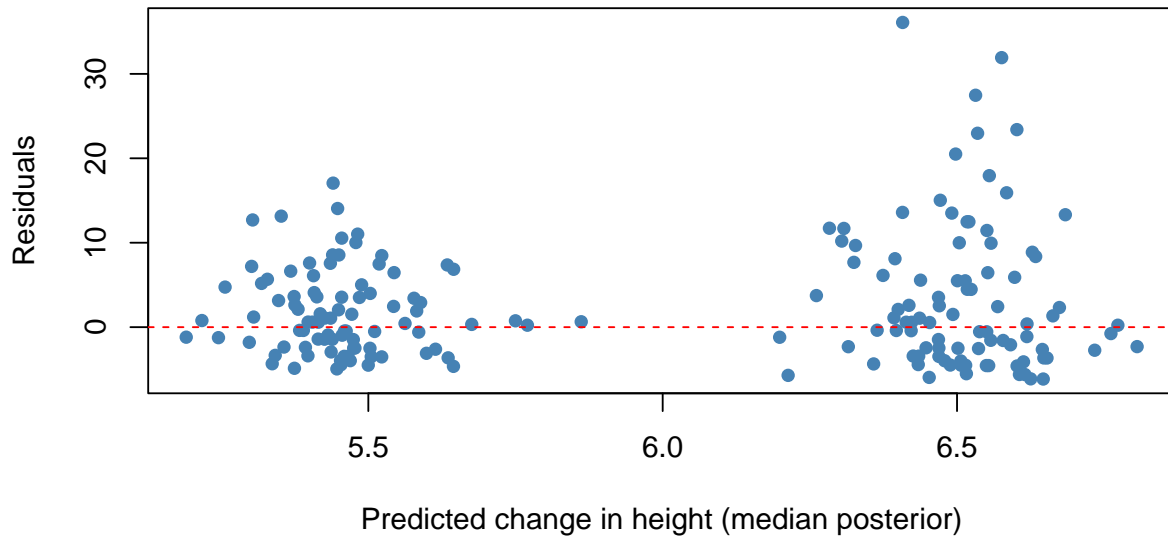
To meet the assumptions of parametric testing, the change in height was transformed using a Box-Cox power of 0.25. After transformation, residuals were approximately normal (Shapiro-Wilk test: Control, $W = 0.297$; Treatment, $W = 0.475$) and homogeneity of variances between treatments was acceptable (F-test, $p = 0.0516$). ANOVA was performed on the transformed values to assess treatment effects. Anova suggests data not significant at $p = 0.1$.

Next trying Bayes GLM. Not expecting different results, but going forward this might be a neater and more understandable approach. because CI and % change are more in-line with the messages we want to give, rather than the weird interpretation of p value etc...

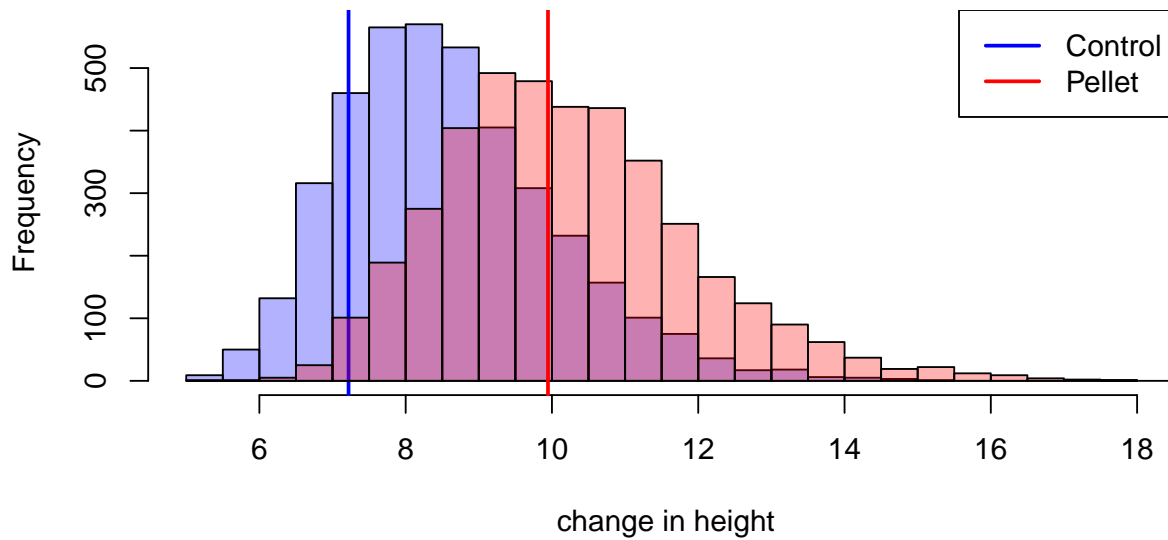
The posterior distribution indicated a median 19% increase in change in height under pellet treatment relative to control, with a $P = 0.80$ probability that the effect was positive. The median 19% increase represents the median of the posterior distribution of percentage differences between pellet and control predictions, obtained by back-transforming posterior samples from the lognormal model. Only positive changes in height were modeled.

Note that these data may be considered unsuitable for these types of models as we have two separate plots (control - Treatment) - this model assumes Independence across trees. (PS we do not need constant variance, balance or normality). That is - it could be said that we have 2 data points - treatment average and control average.

Residuals vs Predicted for Quercus Model



Posterior predictive means



Note that I am not totally sure how to validate these models. I don't think it matters much with the simple model here, because the data is very straightforward. I show a residuals vs fitted, showing 2 clouds - around the 2 means for pellet and treatment - which looks ok? Second I show the actual mean (solid line) with the histogram of predicted model means. Since the actual mean falls within the high density area of the histogram, I think the model is reasonably sensible.

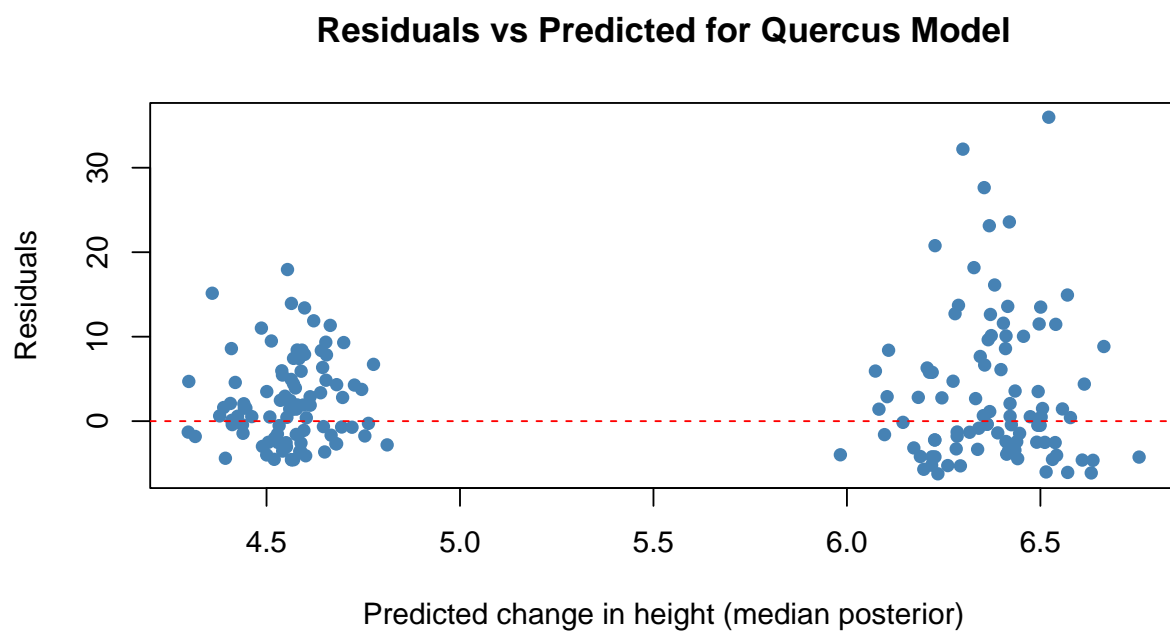
The first table (number) below is the proportion of positive posterior draws on the Treatment change in height, showing the probability of a positive effect - ideally this would be 100%, 90% implies moderate support for positive effect, 80-90% implies weak support.

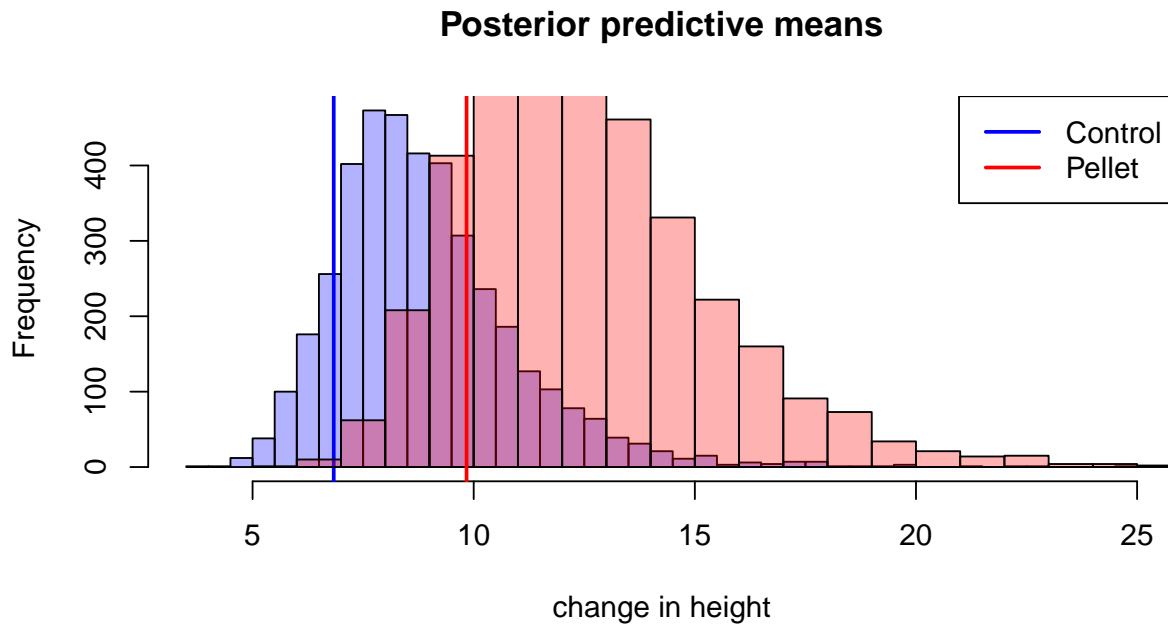
The table shows the posterior median effect and 95% credibility intervals for the pellet applied to oak modeled using a Bayesian linear GLM with lognormal transform to response (delta h per tree).

```
## [1] 0.88125
```

```
## # A tibble: 3 x 4
##   Parameter      Median CI_lower CI_upper
##   <chr>          <dbl>    <dbl>    <dbl>
## 1 Control (Intercept)  5.46      4.48      6.69
## 2 Pellet             6.47      5.33      7.93
## 3 Pellet % change    19.0     -10.4     57.3
```

This second bit of analysis is sensitivity analysis for including the zero values - by adding 0.1 to every delta_h





The validation plot showing the modeled means and the actual mean seem to suggest that the control raw mean is closer to being outside the bulk of the values of the histogram, compared to when zeros were omitted.

The first table/number below is the proportion of positive posterior draws on the Treatment change in height, WHEN WE INCLUDE THE ZEROS but adjust them by 0.1 so that we can log

The summary table shows the model predicted medians, back transformed from lognormal, with 95% credibility intervals

```
## [1] 0.97675
```

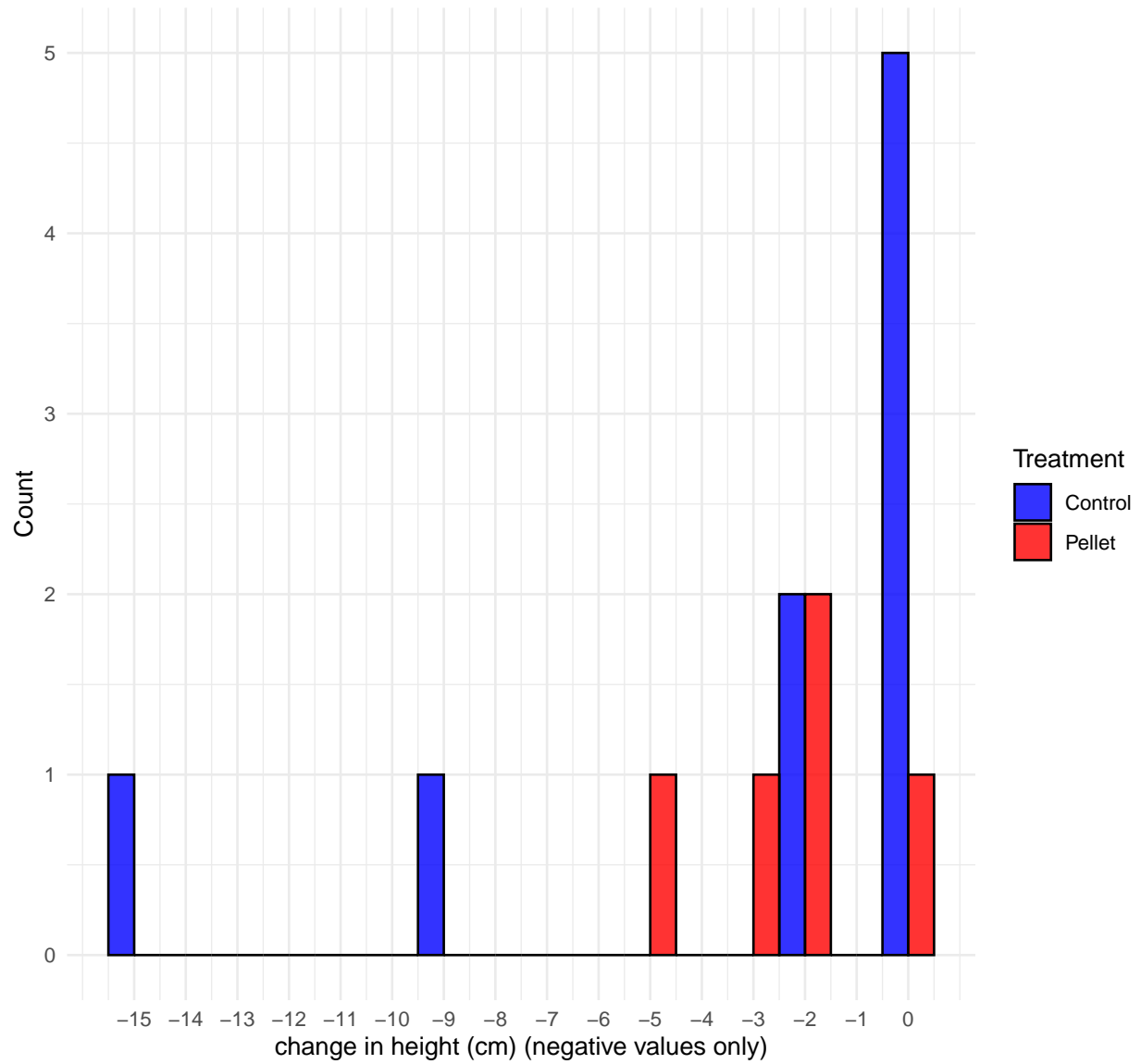
```
## # A tibble: 3 x 4
##   Parameter      Median CI_lower CI_upper
##   <chr>          <dbl>   <dbl>   <dbl>
## 1 Control (Intercept)  4.53    3.57    5.72
## 2 Pellet           6.36    5.02    8.10
## 3 Pellet % change  40.6    0.571   94.1
```

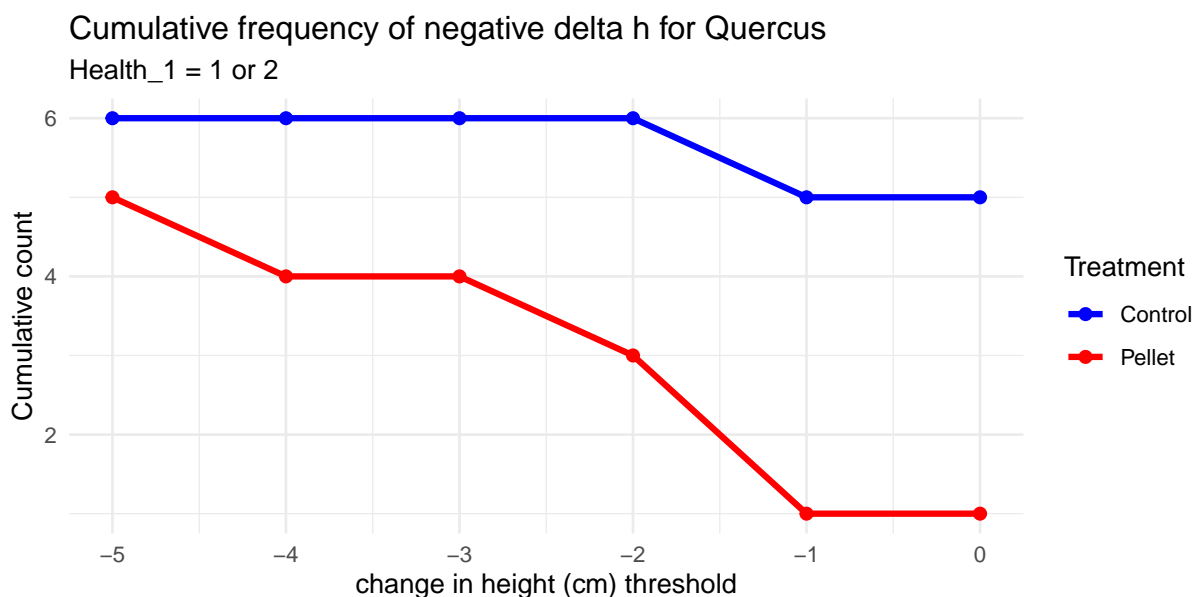
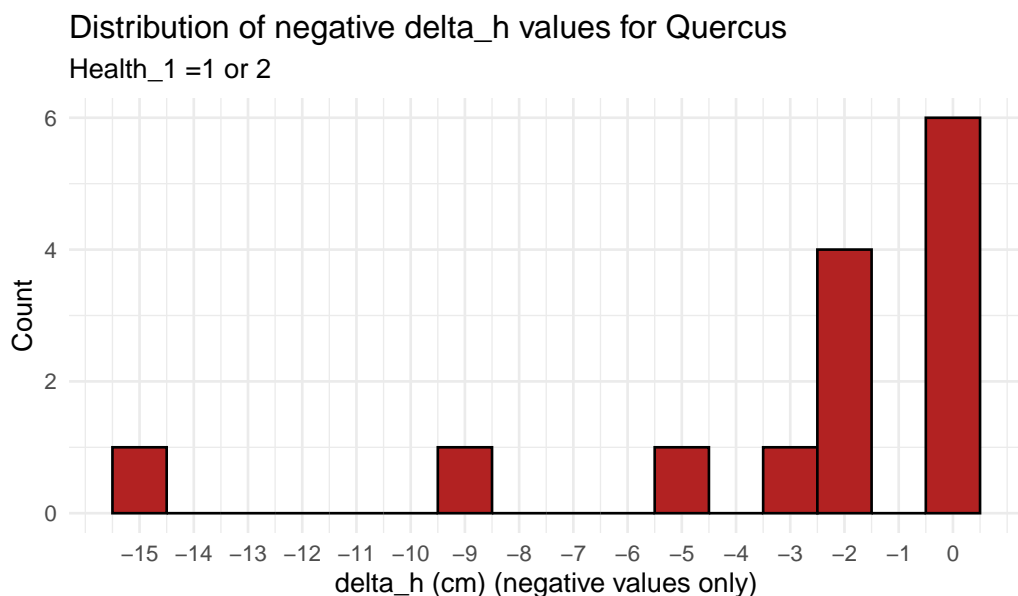
The posterior distribution now indicates a median 40% increase in change in height under pellet treatment relative to control, with a $P = 0.97$ probability that the effect was positive.

Are we getting good result now because we arbitrarily choose 0s to add .1, but what about the -1, -1.5, where do we draw the line. We chose to ignore all -ve - but we know we have ~1cm error, we have measured that. Lets look at distribution of -ve values first, then do some sensitivity analysis.

Distribution of negative delta height values for Quercus

Health_1 = 1 or 2





The cf chart shows that if we leave out 0s, we omit 5 control trees and 1 treatment tree. If we omit anything below -2, we omit 6 control trees and 3 treatment trees., -3, 6 control and 4 treatment.

These data points do carry information - they are most likely to be trees for which change in height was unmeasurably small. That is a result and by omitting them completely we lose that data. But where do we cut off. i) If add back in more -ve control than treatment trees this is in our favour. ii) If we include all trees to -5 we have approximately similar number omitted , but is -5 too much and we are including browsed trees. iii) We have carried out error checks which suggest 1cm is potential error - giving justifiable reason to include -1s.