A Reanalysis of Lung Tumor Incidence in Mice Induced By Inorganic Arsenic Exposure

Lyle D. Burgoon (US Army Engineer Research and Development Center), Ingrid Druwe (ORISE)

December 2, 2015

Introduction

In Cohen et al (2014) and Cohen et al (2015) the authors raised interesting questions with respect to the reproducibility of the control animal tumor incidences in the Tokar et al (2011) and Waalkes et al (2014) studies. Cohen et al's objections, if true, would raise uncertainty about the quality of the Tokar and Waalkes studies. If Cohen et al's objections were true it would be difficult for either study to be used in support of a hazard and dose-response assessment of inorganic arsenic.

We started our analysis by focusing on Cohen et al's central question: whether the control tumor incidences in Waalkes et al differed from that reported in Tokar et al. If there were no substantive difference, then we would use the Tokar control data as a prior, and update the Waalkes data to obtain a more informed posterior distribution. From there, we assessed if there were differences in tumor incidences in the Waalkes data between the controls and inorganic exposed mice.

Analysis

Are Control Tumor Incidences in Tokar et al's Data Different From Waalkes et al's Data?

The Tokar et al (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3003832/ (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3003832/)) and Waalkes et al (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4130362/ (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4130362/)) adenoma and carcinoma data were used.

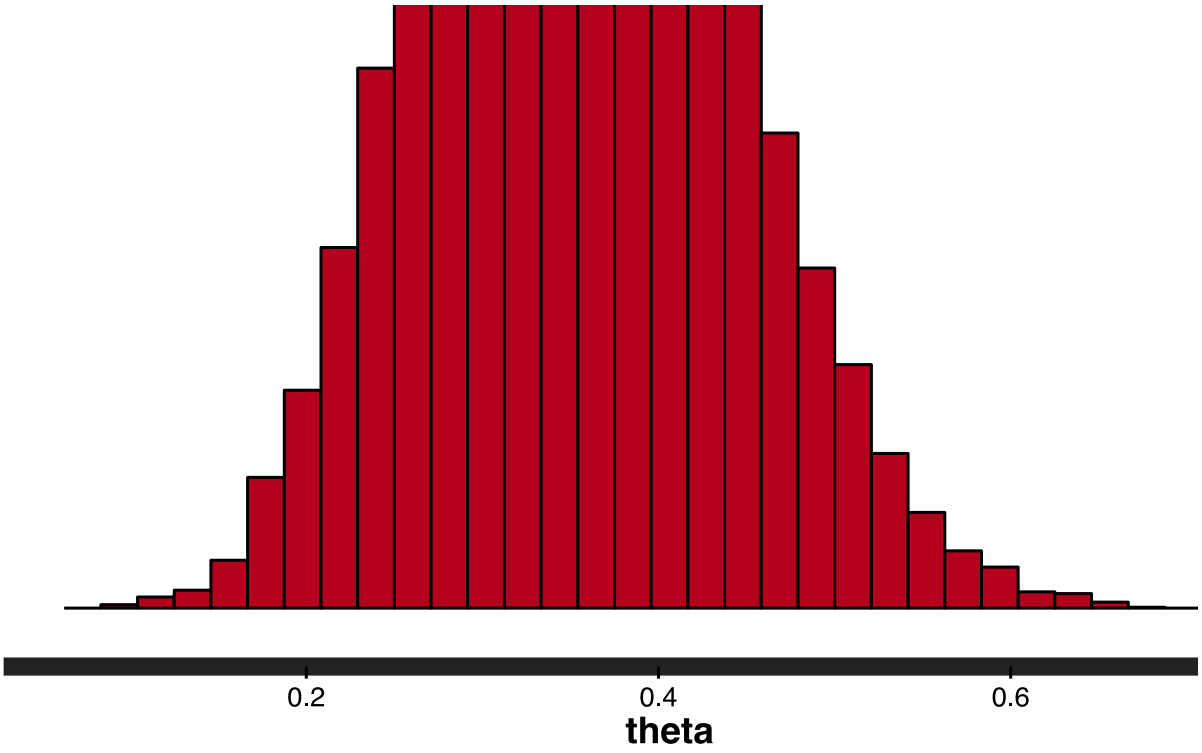
In [2]:
```
library(ggplot2)
library(rstan)
tokar_tumors <- 10
tokar_animals <- 29
waalkes_tumors <- 8
waalkes_animals <- 36
```

We used R and Stan to perform Markov Chain Monte Carlo (MCMC) Bayesian analysis. First, we constructed and compiled the Stan model. Then we used Stan to calculate the posterior probability distribution for the tumor incidences in the controls. We used a Bernoulli distribution to model the data from Tokar et al and Waalkes et al. We used a noninformative beta conjugate distribution as we had no prior knowledge of the control tumor incidences at the NIEHS/NTP laboratories.

In [3]:
```
# THE MODEL.
tokar_modelString = "
data {
  int<lower=0> N;     //number of items
  int y[N];           // y is an N-length vector of ints
}
parameters {
  real <lower=0, upper=1> theta;
}
model {
  theta ~ beta(1,1);
  y ~ bernoulli(theta);
}
"

stanDso <- stan_model( model_code=tokar_modelString )

N <- tokar_animals
z <- tokar_tumors
y <- c( rep( 1, z), rep( 0, N-z))
dataList <- list( y = y , N = N )


#This is going to do the MCMC for Tokar's data. Note we're using a fla
stanFit <- sampling( object = stanDso , data = dataList , chains = 3 ,
stan_hist(stanFit)
```

```
SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002769 seconds (Warm-up)
#                0.049519 seconds (Sampling)
#                0.052288 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
```

```
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002195 seconds (Warm-up)
#                0.050088 seconds (Sampling)
#                0.052283 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002652 seconds (Warm-up)
#                0.046738 seconds (Sampling)
#                0.04939 seconds (Total)


stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```



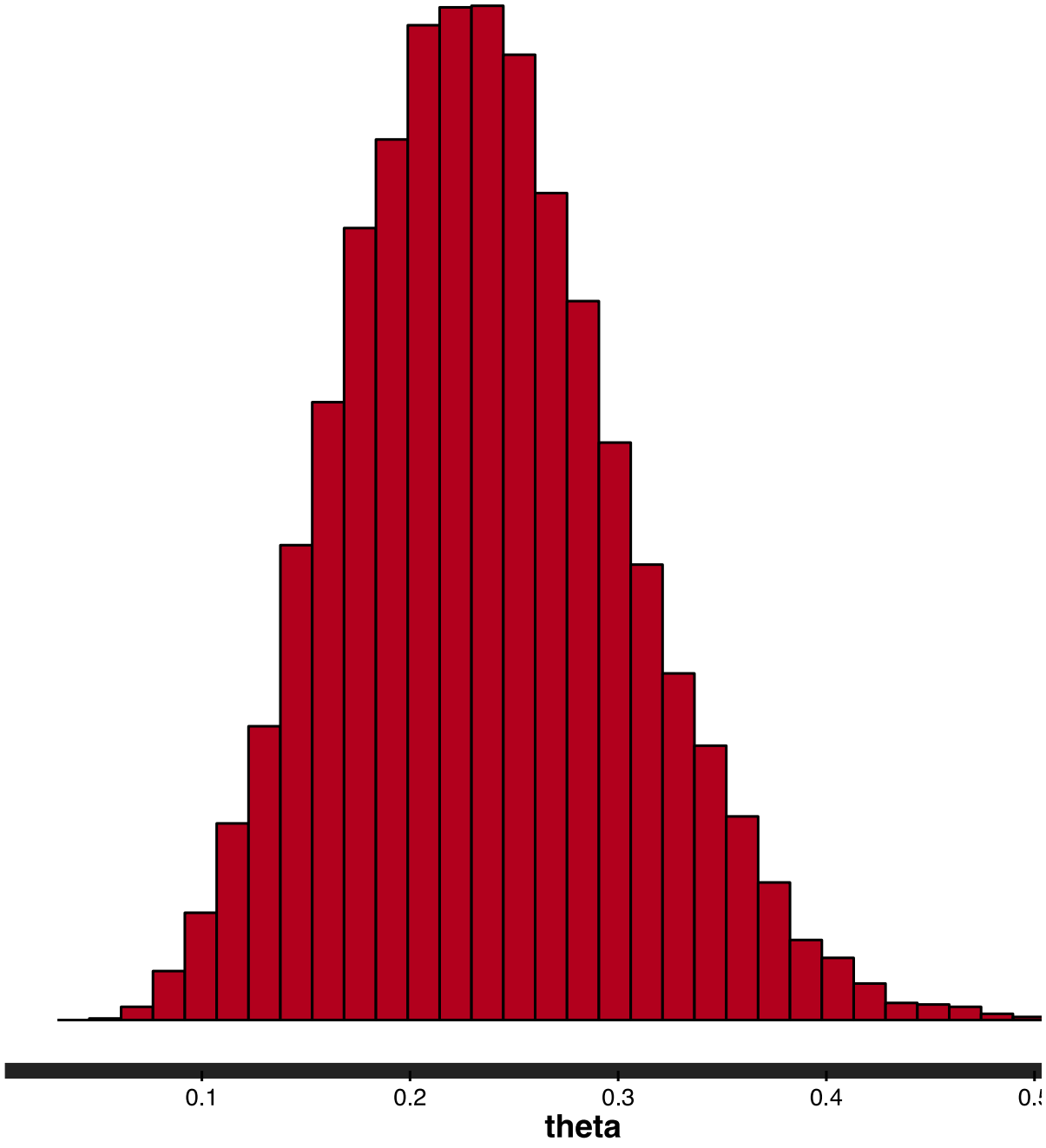
The histogram of posterior probability of the tumor incidences in Tokar's controls shows that it is centered at 35%. In a Bayesian context, the spread of the distribution represents our uncertainty in the control tumor incidences. This distribution is largely informed by by the likelihood from Tokar's data (10/29 = 34.4%).

Next, we are going to analyze Waalkes et al's data. The same model is used. The only difference is that we are using the data from the Waalkes et al paper.

In [4]:
```
Nw <- waalkes_animals
zw <- waalkes_tumors
yw <- c(rep(1,zw), rep(0, Nw-zw))
dataListW <- list(y=yw, N=Nw)
stanFitW <- sampling( object = stanDso , data = dataListW , chains = 3
stan_hist(stanFitW)
```

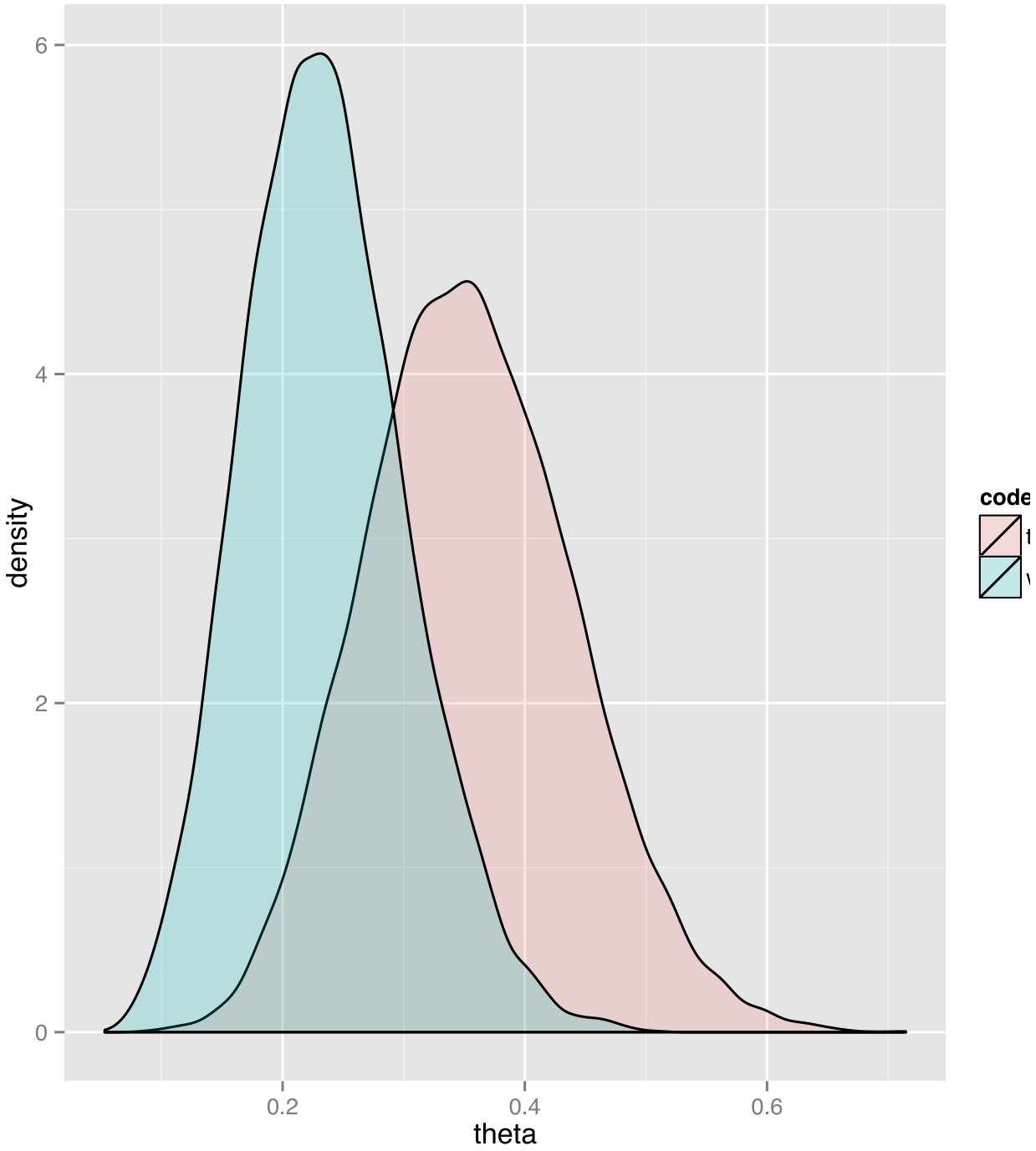
```
SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002286 seconds (Warm-up)
#                0.054905 seconds (Sampling)
#                0.057191 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002281 seconds (Warm-up)
#                0.05258 seconds (Sampling)
#                0.054861 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
```

```
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)

#  Elapsed Time: 0.002026 seconds (Warm-up)
#                0.049577 seconds (Sampling)
#                0.051603 seconds (Total)

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```



The histogram of posterior probability of the tumor incidences in Waalkes et al controls shows that it is centered at 24%. Again, the spread of the distribution represents our uncertainty in the control tumor incidences. This distribution is largely informed by the likelihood from Waalkes et al's data (8/36 = 22.2%).

In [5]:
```
tokar_posterior <- as.data.frame(extract(stanFit)[[1]])
waalkes_posterior <- as.data.frame(extract(stanFitW)[[1]])
colnames(tokar_posterior) <- "theta"
colnames(waalkes_posterior) <- "theta"
combined_data <- rbind(tokar_posterior, waalkes_posterior)
code <- c(rep("tokar", nrow(tokar_posterior)), rep("waalkes", nrow(waa
combined_data <- cbind(combined_data, code=code)
ggplot(combined_data, aes(theta, fill = code)) + geom_density(alpha =
```



Naively, one might conclude that since the distribution centers (medians) are different by about 13% (see above), that the studies have drastically different control tumor incidences. That conclusion would support Cohen et al's argument, and appears to be the basis of

Cohen et al's argument. However, that interpretation ignores several important aspects.

There is a significant amount of uncertainty with respect to the exact control tumor incidences for both Waalkes et al's and Tokar et al's data due to the fact that they are both being sampled from a distribution. As a result, it is arguably possible that both datasets are from the same distribution. Simply looking at the central tendencies or estimates of the two datasets does not allow us to draw a conclusion. However, in looking at the histrograms above, we can see that there is significant overlap between Waalkes et al and Tokar et al control tumor incidence distributions.

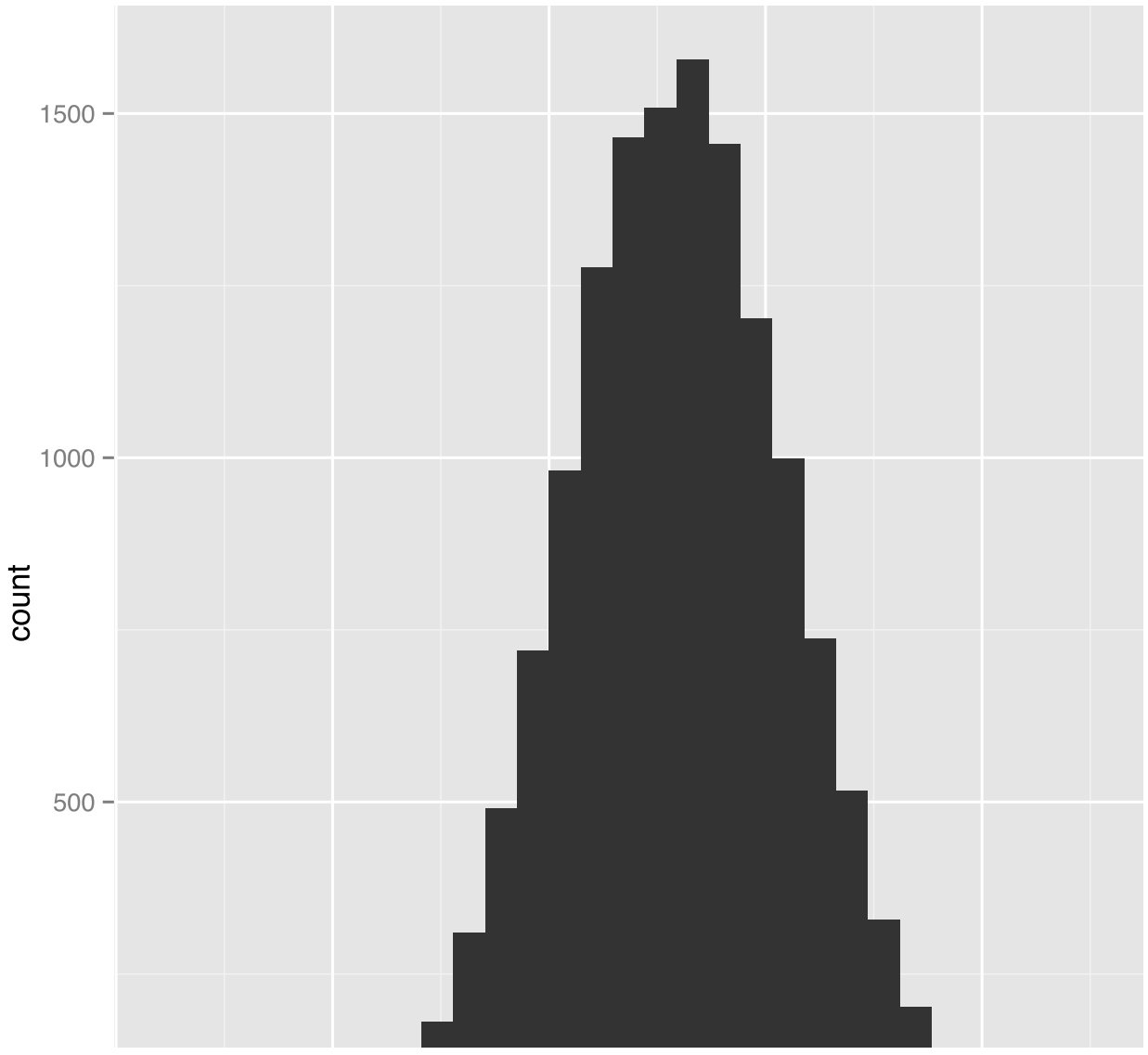
This type of graphical analysis, although intuitive, is not definitive. We need to look at the posterior difference of the draws from both distributions. We are using an approach that sets a region of practical equivalence (ROPE) around the zero difference, and the 95% highest density interval of the difference distribution. The ROPE demarcates a region around zero difference that is functionally equivalent to no difference. The decision rules that we will apply are:

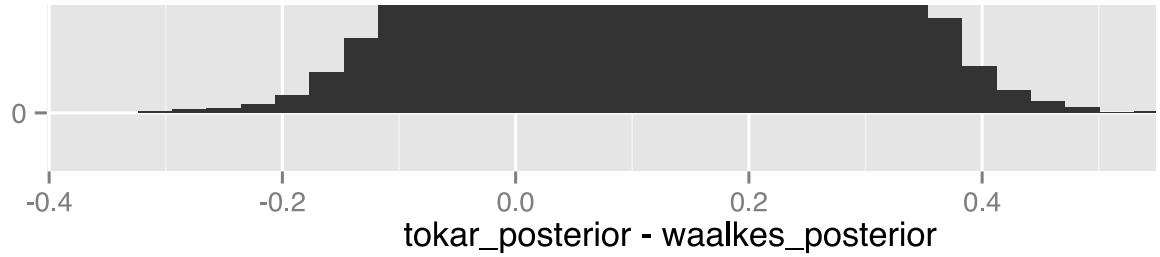
1. If the 95% HDI is completely within the ROPE, then the incedence from the Tokar et al and Waalkes et al studies are the same.
2. If the 95% HDI contains zero, then zero difference is a credible value, meaning the incidences from the Tokar et al and Waalkes et al studies are the same.
3. If the 95% HDI does not contain zero and the 95% HDI is within the ROPE, then we cannot state that the incidences from Tokar et al and Waalkes et al are different, nor can we say strongly that they are the same. More data is required.
4. If the 95% HDI is completely outside the ROPE, then the incidences from the Tokar et al and Waalkes et al studies are different.

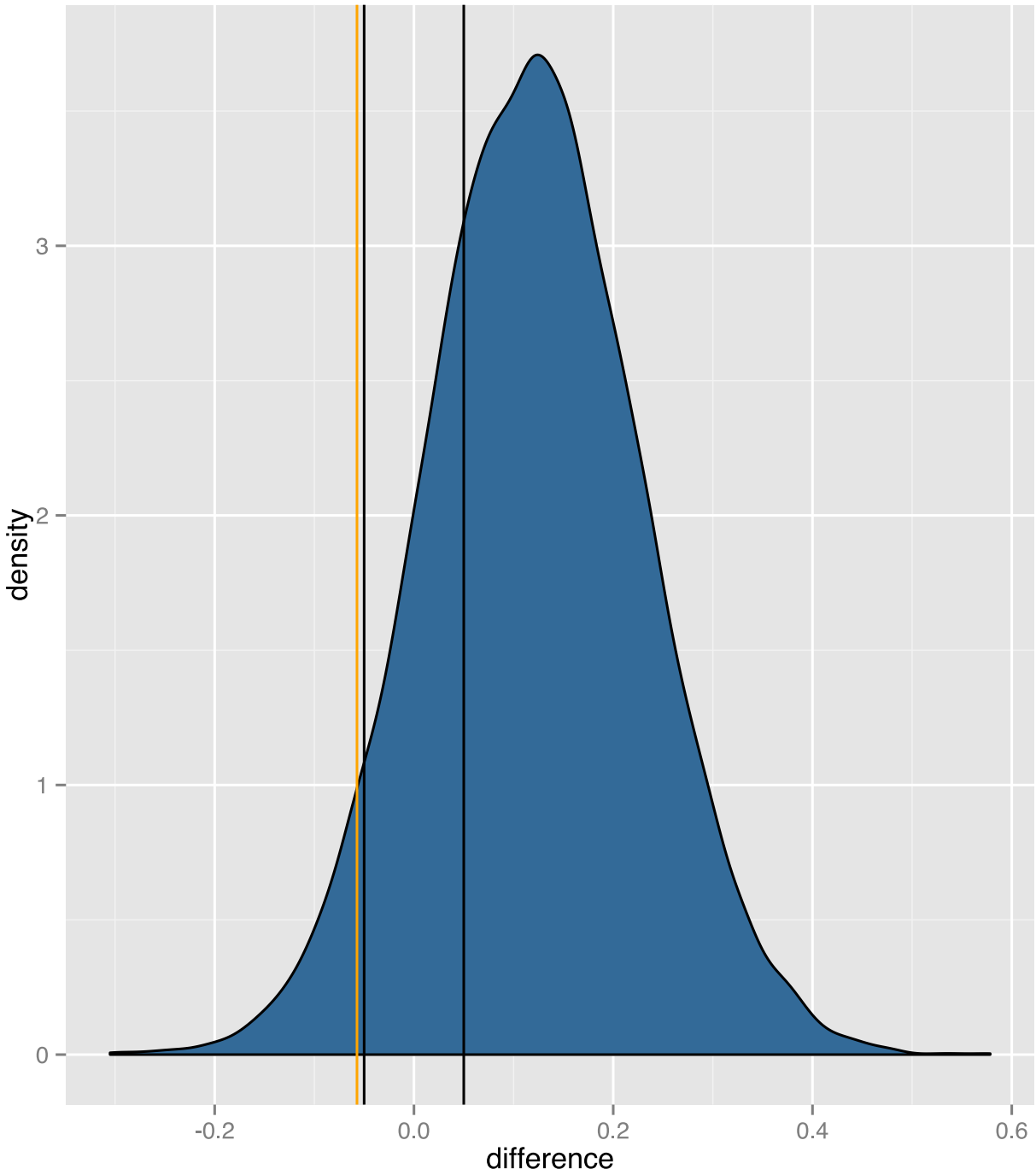
In [6]:
```
tokar_posterior <- extract(stanFit)[[1]]
waalkes_posterior <- extract(stanFitW)[[1]]
qplot(tokar_posterior - waalkes_posterior, geom="histogram")
diff_distro <- as.data.frame(tokar_posterior - waalkes_posterior)
colnames(diff_distro) <- "difference"

#How much of the distribution is within the ROPE?
diff_ecdf <- ecdf(diff_distro[,1])
diff_ecdf(.05) - diff_ecdf(-.05)
#21%

ggplot(diff_distro, aes(difference, fill=1)) +
  geom_density(alpha = 1) +
  geom_vline(xintercept = -0.05) +
  geom_vline(xintercept = 0.05) +
  geom_vline(xintercept = quantile(diff_distro[,1], probs=c(0.05)), co
```

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```

Out[6]: 0.202222222222222







We used a ROPE of +/- 0.05 centered at a difference of zero (the black vertical lines on the histogram demarcate the boundaries of the ROPE). We applied our decision rules and concluded that the incidences from both studies are likely from the same distribution, and

thus are likely the same. The difference of 13% between the incidences is likely a manifestation of a small sampling from the same distribution. Note that the posterior difference distribution also includes 0 difference within the 95% HDI (all of the distribution that is greater than the orange line), and that 21% of the posterior difference distribution is within the ROPE.

Does inorganic arsenic exposure increase the tumor incidence in the Waalkes et al study?

Based on our analysis we have established that Cohen et al's argument that the control tumor incidences are different in the two studies is invalid. Next, we assessed if inorganic arsenic exposure increased the tumor incidence in the Waalkes et al study at different doses. To do this, we used Tokar et al's control data as the prior, and Waalkes et al's control data as the likelihood and constructed the posterior distribution for the control tumor incidence across the two studies. Then we compared the posterior distribution control tumor incidence to the posterior distribution for the tumor incidences in the various inorganic arsenic exposure levels. We used a flat uninformative prior to generate the posterior distributions for each arsenic exposure. Then we sampled from the posterior distribution for each arsenic exposure and we sampled from the posterior distribution for the control tumor incidence. Finally, we took the difference of these values and constructed a posterior distribution of the difference between each arsenic exposure concentration and the controls.

Construct Posterior Probability for Control Tumor Incidence

First, we constructed the control tumor incidence posterior distribution. We had to build a new model. We had already analytically solved to identify that Tokar et al's control tumor incidence is properly modeled with by a beta(10, 19) distribution.

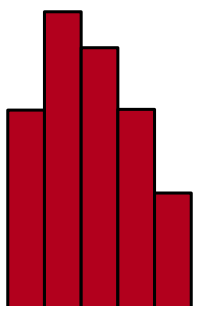
In [7]:
```
# THE MODEL.
control_tumor_modelString = "
data {
  int<lower=0> N;     //number of items
  int y[N];           // y is an N-length vector of ints
}
parameters {
  real <lower=0, upper=1> theta;
}
model {
  theta ~ beta(10, 19); //prior distribution
  y ~ bernoulli(theta); //posterior
}
"
control_tumor_stanDso <- stan_model( model_code=control_tumor_modelStr
control_tumor_stanFit <- sampling( object = control_tumor_stanDso , da
control_tumor_stanFit
stan_hist(control_tumor_stanFit)
```

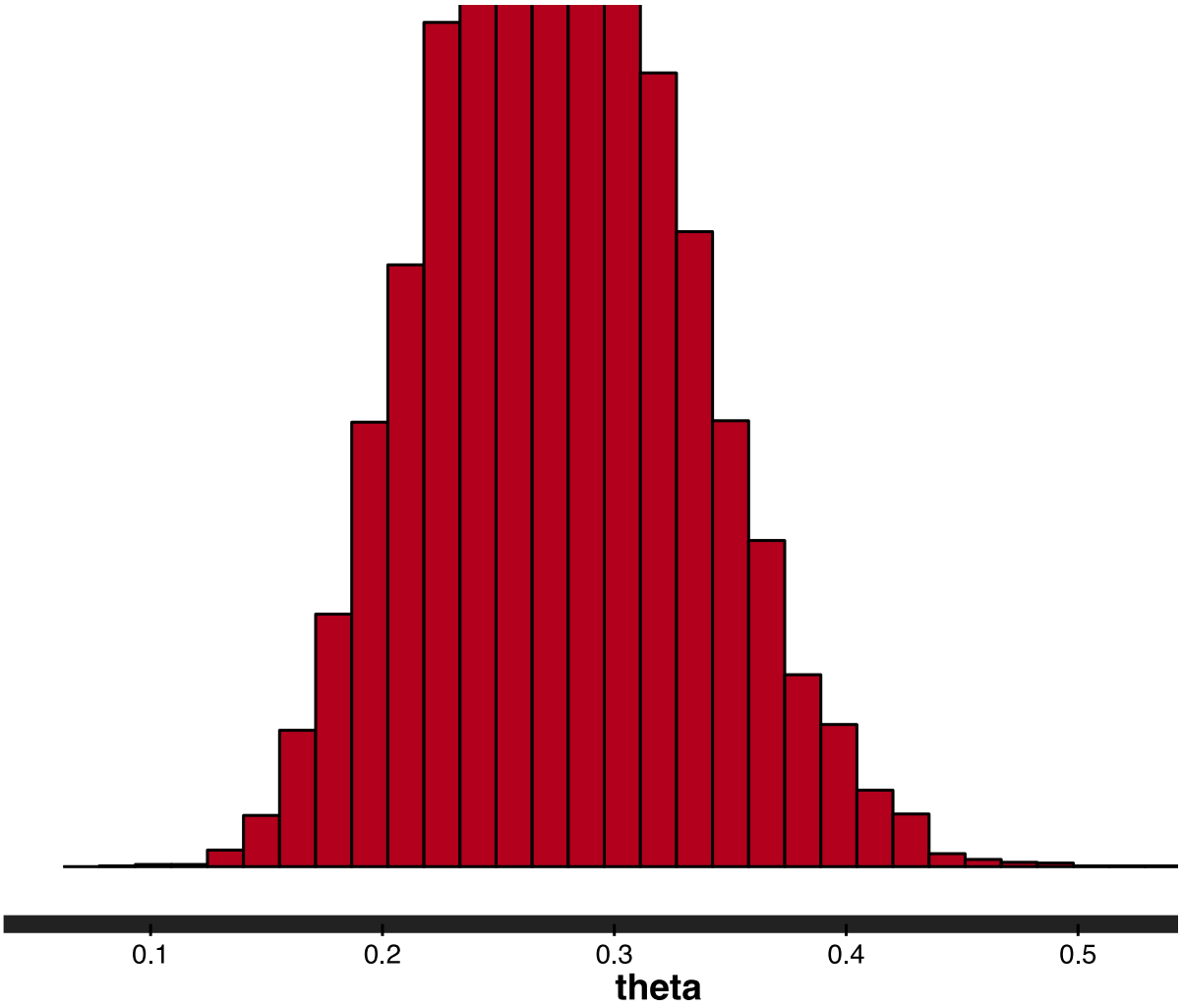
```
SAMPLING FOR MODEL '0b6190a93989eddbfcf8a45ab8d2c543' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002223 seconds (Warm-up)
#                0.045373 seconds (Sampling)
#                0.047596 seconds (Total)


SAMPLING FOR MODEL '0b6190a93989eddbfcf8a45ab8d2c543' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
```

```
#  Elapsed Time: 0.002415 seconds (Warm-up)
#                0.046964 seconds (Sampling)
#                0.049379 seconds (Total)


SAMPLING FOR MODEL '0b6190a93989eddbfcf8a45ab8d2c543' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.00209 seconds (Warm-up)
#                0.042877 seconds (Sampling)
#                0.044967 seconds (Total)
```

Out[7]:

```
Inference for Stan model: 0b6190a93989eddbfcf8a45ab8d2c543.
3 chains, each with iter=5000; warmup=200; thin=1;
post-warmup draws per chain=4800, total post-warmup draws=14400.

        mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff R
hat
theta   0.28    0.00 0.06   0.17   0.24   0.27   0.31   0.39  5459
1
lp__  -38.87    0.01 0.73 -40.89 -39.03 -38.59 -38.41 -38.35  4621
1

Samples were drawn using NUTS(diag_e) at Thu Dec 17 13:33:17 2015.
For each parameter, n_eff is a crude measure of effective sample siz
e,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```





We can see that the distribution has a mean of 28%. Thus, we would estimate that the overall control tumor incidence is approximately 28%, after taking into consideration our prior knowledge from Tokar et al and the latest data from Waalkes et al.

50ppb inorganic arsenic

We will assess the Waalkes et al tumor incidences in mice exposed to inorganic arsenic in the same way as we assessed if there were differences between the control tumor incidences in the Tokar et al and Waalkes et al studies. The only difference is that we used the updated posterior distribution for control tumor incidences we just calculated above.

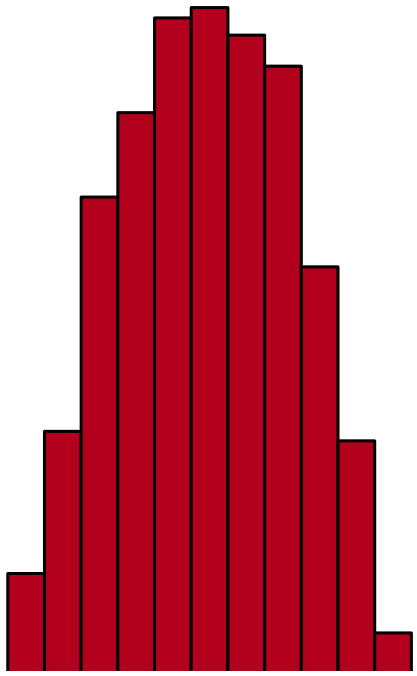
In [8]:
```
# 50ppb inorganic arsenic
Nw <- 37  #total animals
zw <- 19  #total tumors -- adenomas or carcinomas
yw <- c(rep(1,zw), rep(0, Nw-zw))
dataList50ppb <- list(y=yw, N=Nw)
stanFit50ppb <- sampling( object = stanDso , data = dataList50ppb , ch
stanFit50ppb
stan_hist(stanFit50ppb)
```

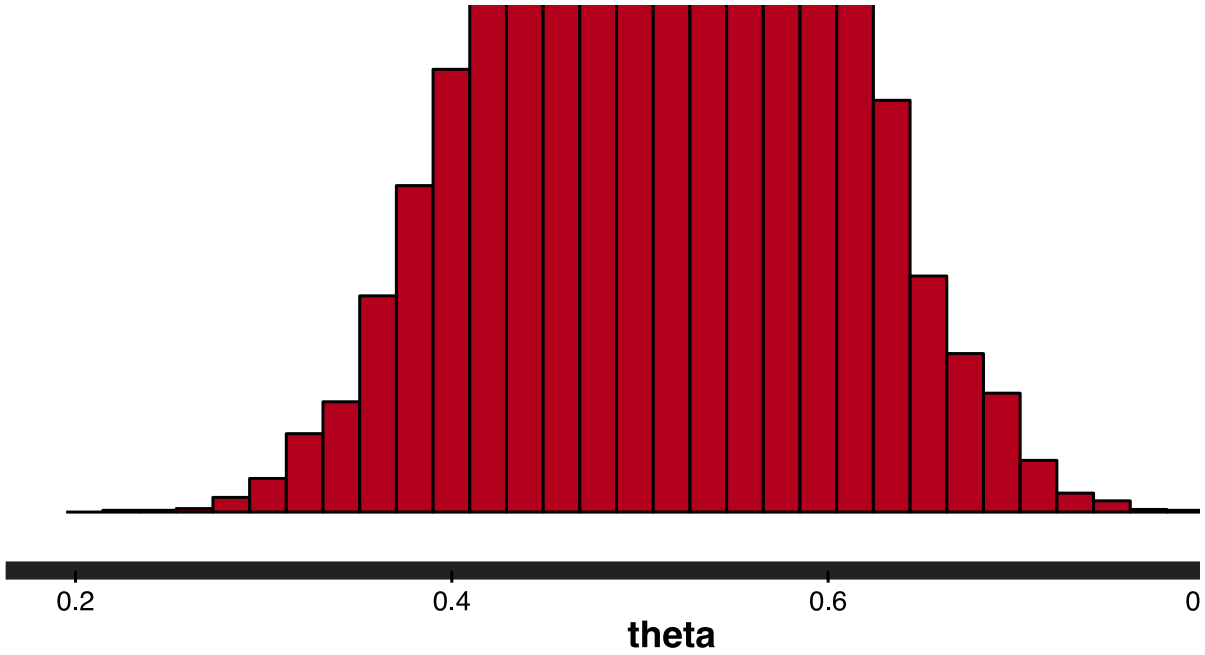
```
SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002147 seconds (Warm-up)
#                0.05116 seconds (Sampling)
#                0.053307 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.00205 seconds (Warm-up)
#                0.049021 seconds (Sampling)
#                0.051071 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
```

```
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002259 seconds (Warm-up)
#                0.050079 seconds (Sampling)
#                0.052338 seconds (Total)
```

Out[8]:

```
Inference for Stan model: 2dcba63f61ddd7c9bd68c7381443898a.
3 chains, each with iter=5000; warmup=200; thin=1;
post-warmup draws per chain=4800, total post-warmup draws=14400.

        mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff R
hat
theta   0.51    0.00 0.08   0.36   0.46   0.51   0.57   0.67  5662
1
lp__  -27.53    0.01 0.72 -29.63 -27.70 -27.25 -27.07 -27.02  4359
1

Samples were drawn using NUTS(diag_e) at Thu Dec 17 13:33:18 2015.
For each parameter, n_eff is a crude measure of effective sample siz
e,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```



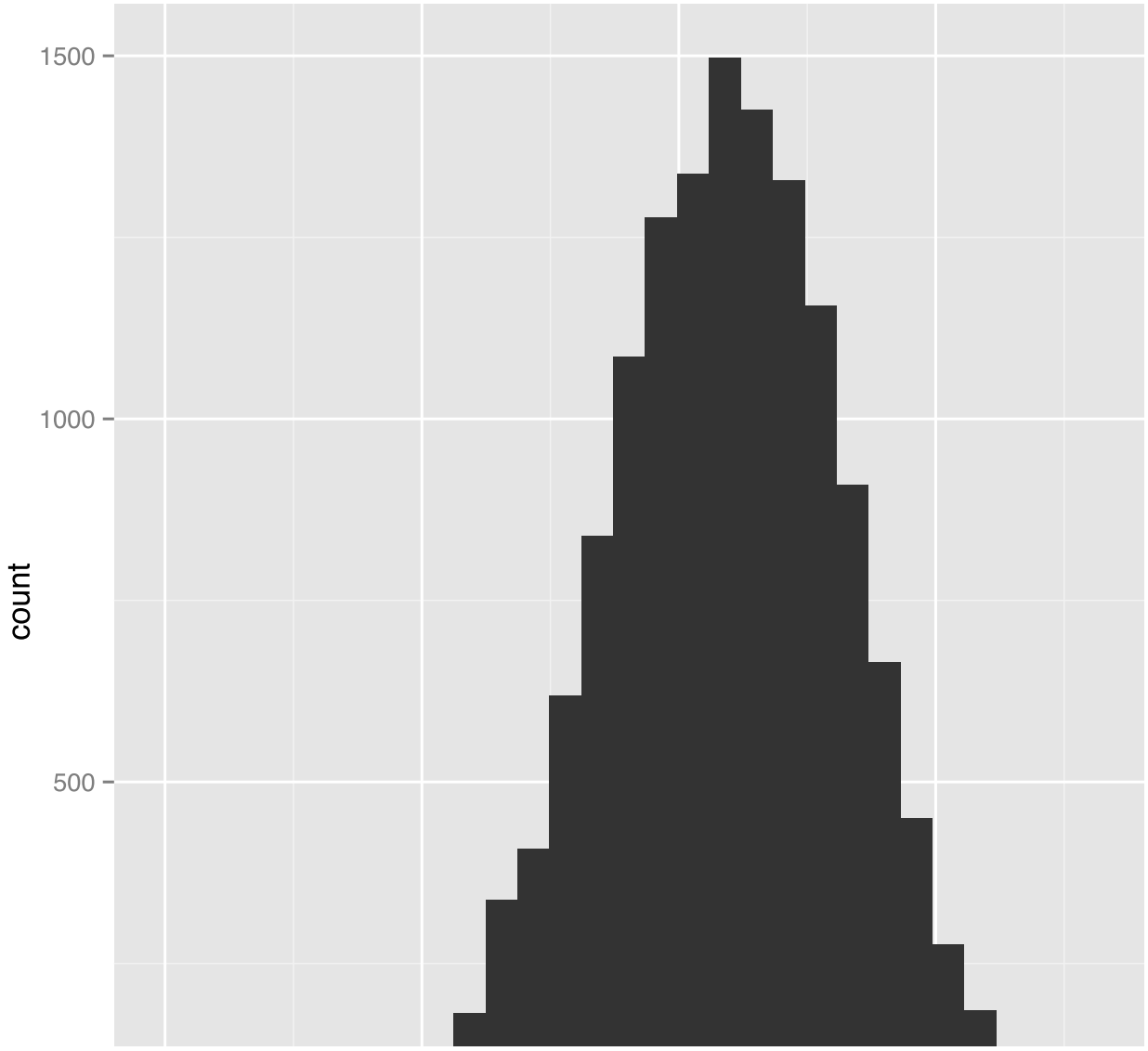


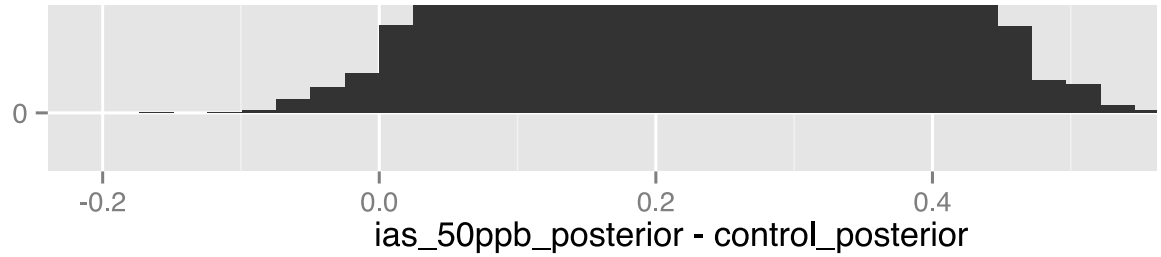
Based on this analysis, the average tumor incidence in the 50ppb exposed group is 51%. This is similar to the result seen in Waalkes et al. Next we will perform the difference analysis.

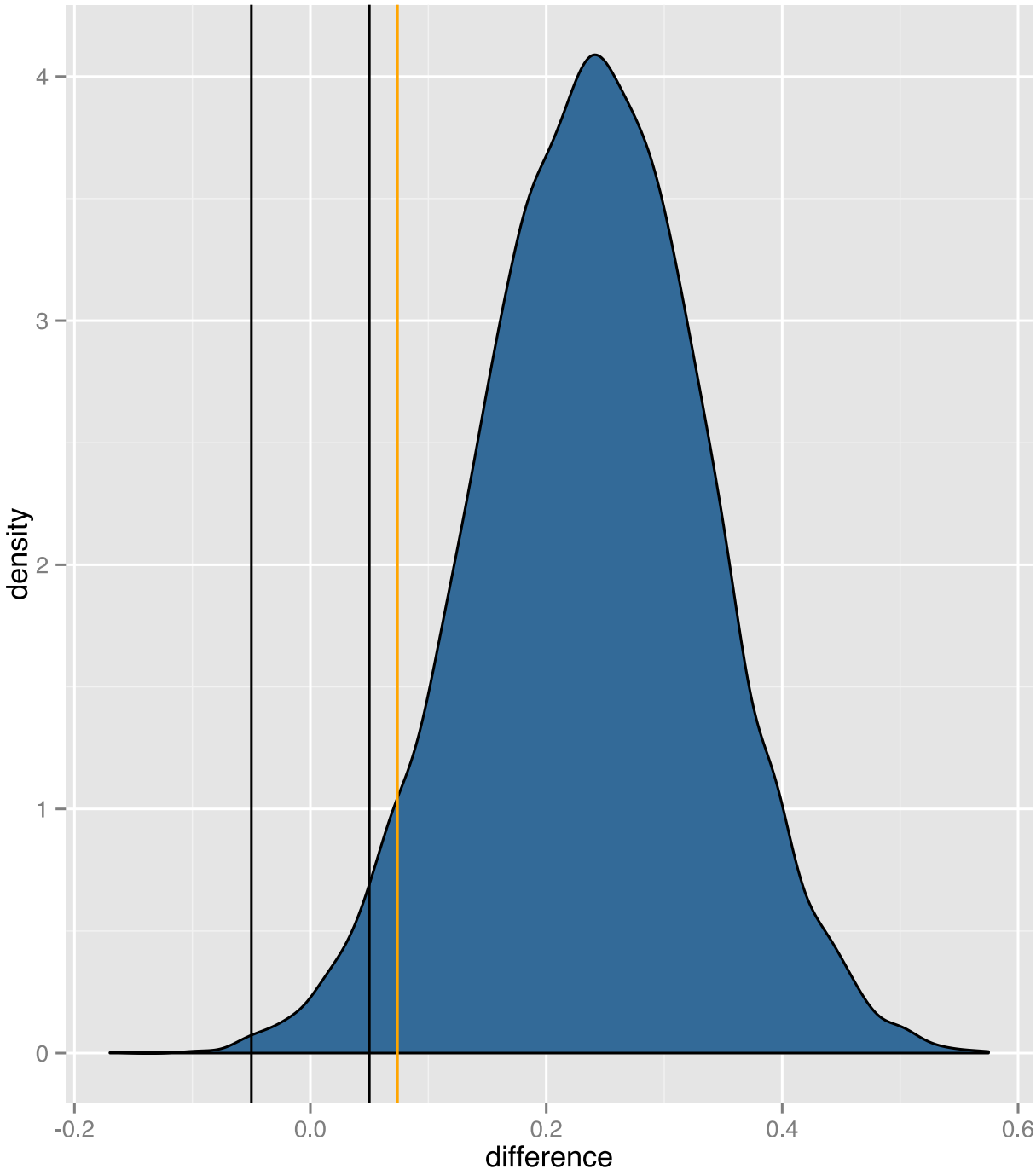
In [9]:
```
control_posterior <- extract(control_tumor_stanFit)[[1]]
ias_50ppb_posterior <- extract(stanFit50ppb)[[1]]
qplot(ias_50ppb_posterior - control_posterior, geom="histogram")
diff_distro <- as.data.frame(ias_50ppb_posterior - control_posterior)
colnames(diff_distro) <- "difference"

#How much of the distribution is within the ROPE?
diff_ecdf <- ecdf(diff_distro[,1])
diff_ecdf(.05) - diff_ecdf(-.05)
#0.03%

ggplot(diff_distro, aes(difference, fill=1)) +
  geom_density(alpha = 1) +
  geom_vline(xintercept = -0.05) +
  geom_vline(xintercept = 0.05) +
  geom_vline(xintercept = quantile(diff_distro[,1], probs=c(0.05)), co
```

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```

Out[9]: 0.0263194444444444







The 95% HDI was completely outside of the ROPE, thus the null hypothesis was rejected. However, expanding the ROPE slightly, such as to +/-0.10 would bring the 95% HDI within the ROPE. We do not feel that a +/-10% rope is justified, when dealing with tumor incidences

in the control population that are approximately 28%. Specifically, we do not believe it is biologically plausible that a tumor incidence that spans from 18-38% can be considered practically equivalent, whereas we feel that one that spans from 23-33% is. Thus, based on our decision criteria, we find that 50ppb inorganic arsenic exposure causes an increase in the lung adenoma and carcinoma tumor incidence in mice.

In [10]:
```
#odds ratio
odds_exposed_tumor <- .51/(1-.51)
odds_non_exposed_tumor <- .28/(1-.28)
odds_ratio <- odds_exposed_tumor / odds_non_exposed_tumor
odds_ratio
```

Out[10]: 2.67638483965015

Said another way, 2.7:1 are the odds of a mouse exposed to 50ppb inorganic arsenic developing lung tumors compared to a non-exposed mouse

500ppb inorganic arsenic

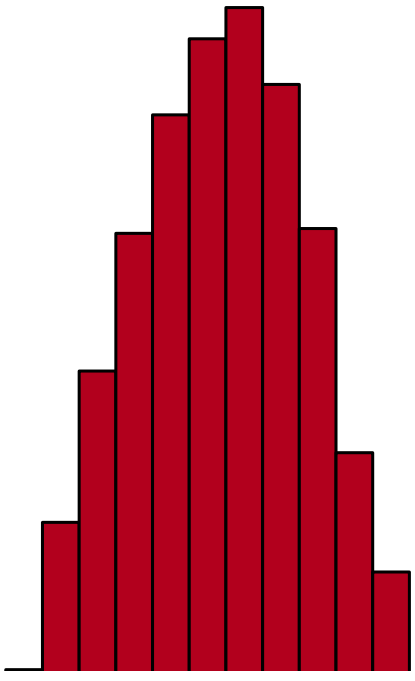
In [11]:
```
# 500ppb inorganic arsenic
Nw <- 37  #total animals
zw <- 20  #total tumors -- adenomas or carcinomas
yw <- c(rep(1,zw), rep(0, Nw-zw))
dataList500ppb <- list(y=yw, N=Nw)
stanFit500ppb <- sampling( object = stanDso , data = dataList500ppb ,
stanFit500ppb
stan_hist(stanFit500ppb)
```

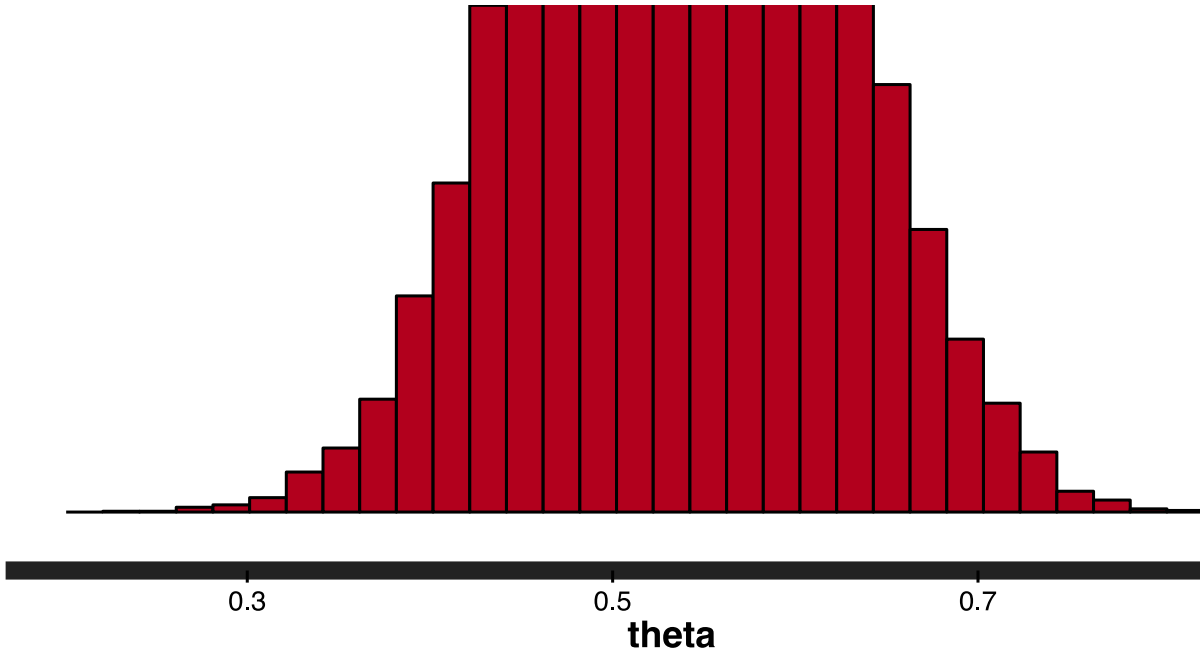
```
SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002358 seconds (Warm-up)
#                0.051822 seconds (Sampling)
#                0.05418 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002696 seconds (Warm-up)
#                0.048376 seconds (Sampling)
#                0.051072 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
```

```
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002446 seconds (Warm-up)
#                0.049208 seconds (Sampling)
#                0.051654 seconds (Total)
```

Out[11]:

```
Inference for Stan model: 2dcba63f61ddd7c9bd68c7381443898a.
3 chains, each with iter=5000; warmup=200; thin=1;
post-warmup draws per chain=4800, total post-warmup draws=14400.

        mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff R
hat
theta   0.54    0.00 0.08   0.39   0.48   0.54   0.59   0.69  4260
1
lp__  -27.42    0.01 0.72 -29.44 -27.59 -27.15 -26.97 -26.92  5332
1

Samples were drawn using NUTS(diag_e) at Thu Dec 17 13:33:19 2015.
For each parameter, n_eff is a crude measure of effective sample siz
e,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```



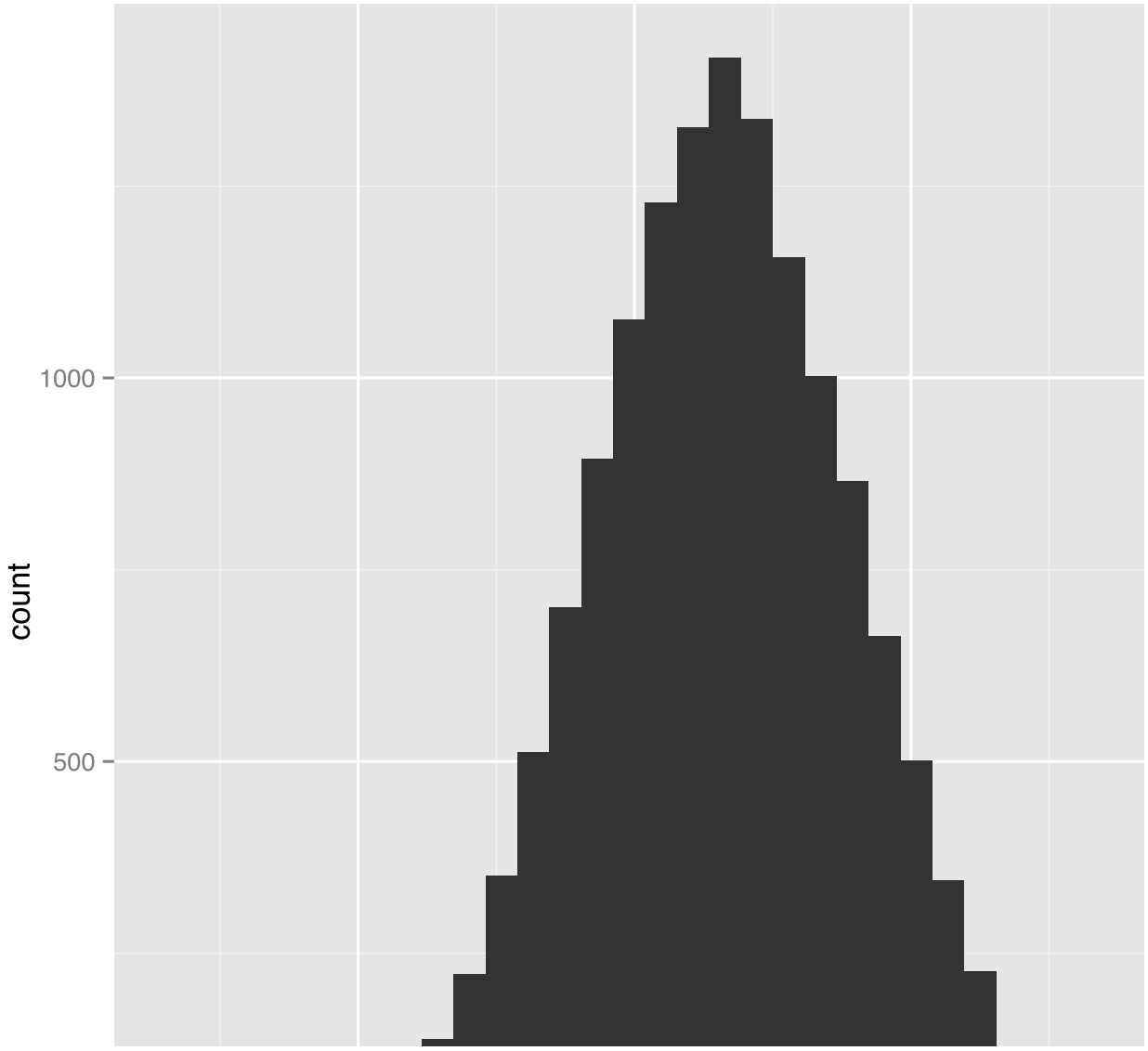


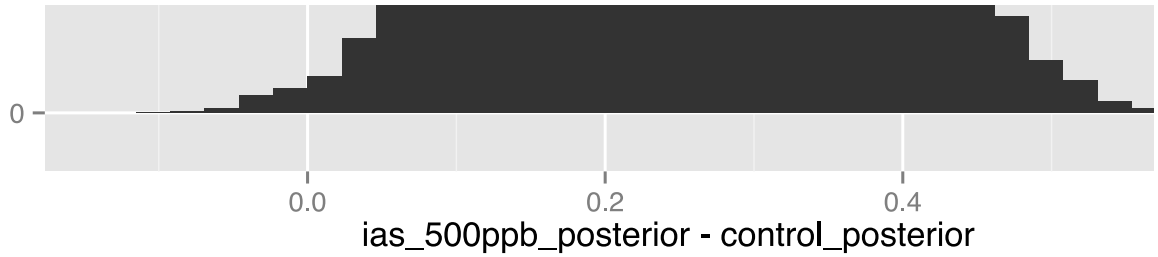
Based on this analysis, the average tumor incidence in the 500ppb exposed group is 54%. This is similar to the result in Waalkes et al. Next, we performed the difference analysis.

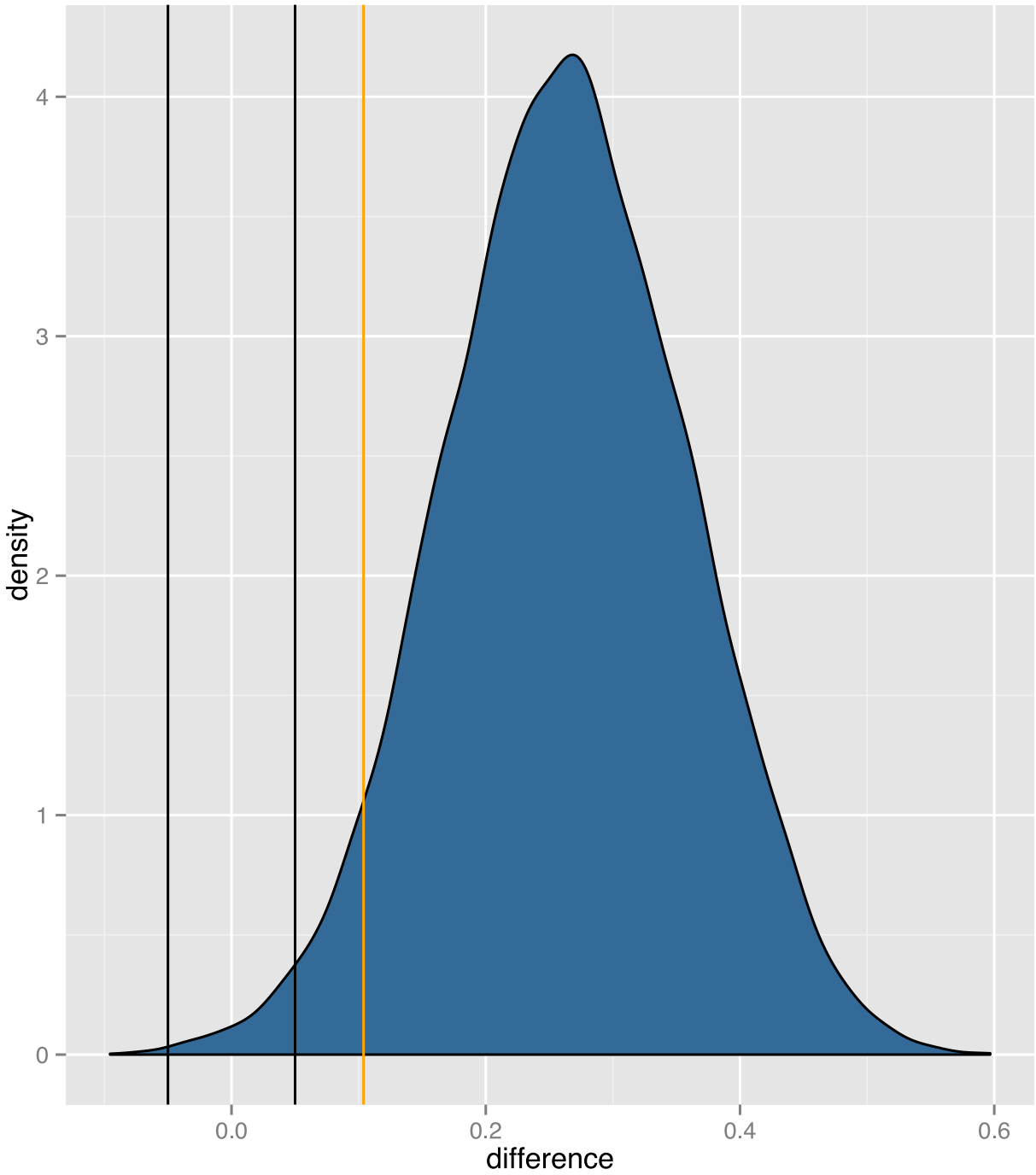
In [12]:
```
control_posterior <- extract(control_tumor_stanFit)[[1]]
ias_500ppb_posterior <- extract(stanFit500ppb)[[1]]
qplot(ias_500ppb_posterior - control_posterior, geom="histogram")
diff_distro <- as.data.frame(ias_500ppb_posterior - control_posterior)
colnames(diff_distro) <- "difference"

#How much of the distribution is within the ROPE?
diff_ecdf <- ecdf(diff_distro[,1])
diff_ecdf(.05) - diff_ecdf(-.05)
#0.02%

ggplot(diff_distro, aes(difference, fill=1)) +
  geom_density(alpha = 1) +
  geom_vline(xintercept = -0.05) +
  geom_vline(xintercept = 0.05) +
  geom_vline(xintercept = quantile(diff_distro[,1], probs=c(0.05)), co
```

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```

Out[12]: 0.0143055555555556







The 95% HDI is completely outside of the ROPE, thus the null hypothesis is rejected. Thus, based on our decision criteria, we find that 500ppb inorganic arsenic exposure causes an increase in the lung adenoma and carcinoma tumor incidence in mice.

In [13]:
```
#odds ratio
odds_exposed_tumor <- .54/(1-.54)
odds_non_exposed_tumor <- .28/(1-.28)
odds_ratio <- odds_exposed_tumor / odds_non_exposed_tumor
odds_ratio
```

Out[13]: 3.01863354037267

Said another way, 3.0:1 were the odds that a mouse exposed to 500ppb inorganic arsenic would develop lung tumors compared to a non-exposed mouse.

5000ppb inorganic arsenic

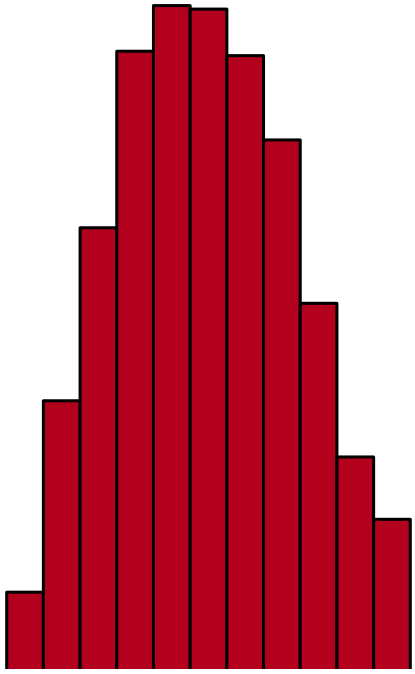
In [14]:
```
# 5000ppb inorganic arsenic
Nw <- 39  #total animals
zw <- 11  #total tumors -- adenomas or carcinomas
yw <- c(rep(1,zw), rep(0, Nw-zw))
dataList5000ppb <- list(y=yw, N=Nw)
stanFit5000ppb <- sampling( object = stanDso , data = dataList5000ppb
stanFit5000ppb
stan_hist(stanFit5000ppb)
```

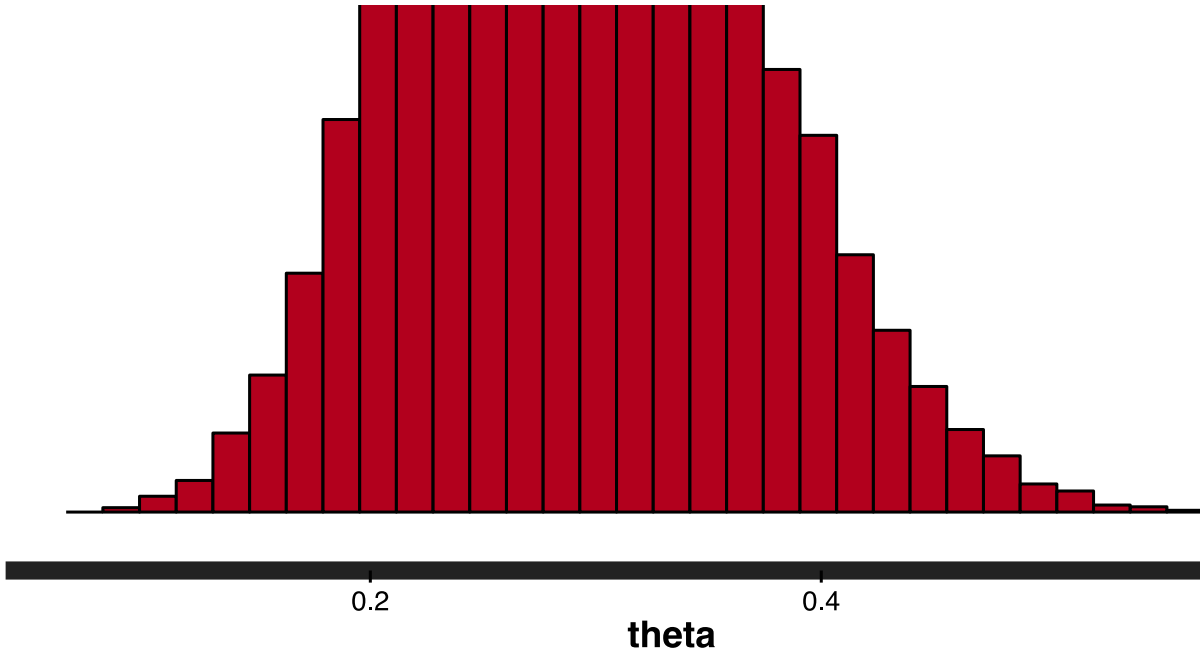
```
SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 1).

Chain 1, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 1, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 1, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 1, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 1, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 1, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 1, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 1, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 1, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 1, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 1, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 1, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002374 seconds (Warm-up)
#                0.04839 seconds (Sampling)
#                0.050764 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 2).

Chain 2, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 2, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 2, Iteration:  700 / 5000 [ 14%]  (Sampling)
Chain 2, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 2, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 2, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 2, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 2, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 2, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 2, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 2, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 2, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.002681 seconds (Warm-up)
#                0.051493 seconds (Sampling)
#                0.054174 seconds (Total)


SAMPLING FOR MODEL '2dcba63f61ddd7c9bd68c7381443898a' NOW (CHAIN 3).

Chain 3, Iteration:    1 / 5000 [  0%]  (Warmup)
Chain 3, Iteration:  201 / 5000 [  4%]  (Sampling)
Chain 3, Iteration:  700 / 5000 [ 14%]  (Sampling)
```

```
Chain 3, Iteration: 1200 / 5000 [ 24%]  (Sampling)
Chain 3, Iteration: 1700 / 5000 [ 34%]  (Sampling)
Chain 3, Iteration: 2200 / 5000 [ 44%]  (Sampling)
Chain 3, Iteration: 2700 / 5000 [ 54%]  (Sampling)
Chain 3, Iteration: 3200 / 5000 [ 64%]  (Sampling)
Chain 3, Iteration: 3700 / 5000 [ 74%]  (Sampling)
Chain 3, Iteration: 4200 / 5000 [ 84%]  (Sampling)
Chain 3, Iteration: 4700 / 5000 [ 94%]  (Sampling)
Chain 3, Iteration: 5000 / 5000 [100%]  (Sampling)
#  Elapsed Time: 0.00228 seconds (Warm-up)
#                0.053411 seconds (Sampling)
#                0.055691 seconds (Total)
```

Out[14]:

```
Inference for Stan model: 2dcba63f61ddd7c9bd68c7381443898a.
3 chains, each with iter=5000; warmup=200; thin=1;
post-warmup draws per chain=4800, total post-warmup draws=14400.

        mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff R
hat
theta   0.29    0.00 0.07   0.17   0.24   0.29   0.34   0.44  5450
1
lp__  -25.29    0.01 0.72 -27.35 -25.44 -25.01 -24.84 -24.79  4709
1

Samples were drawn using NUTS(diag_e) at Thu Dec 17 13:33:20 2015.
For each parameter, n_eff is a crude measure of effective sample siz
e,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```



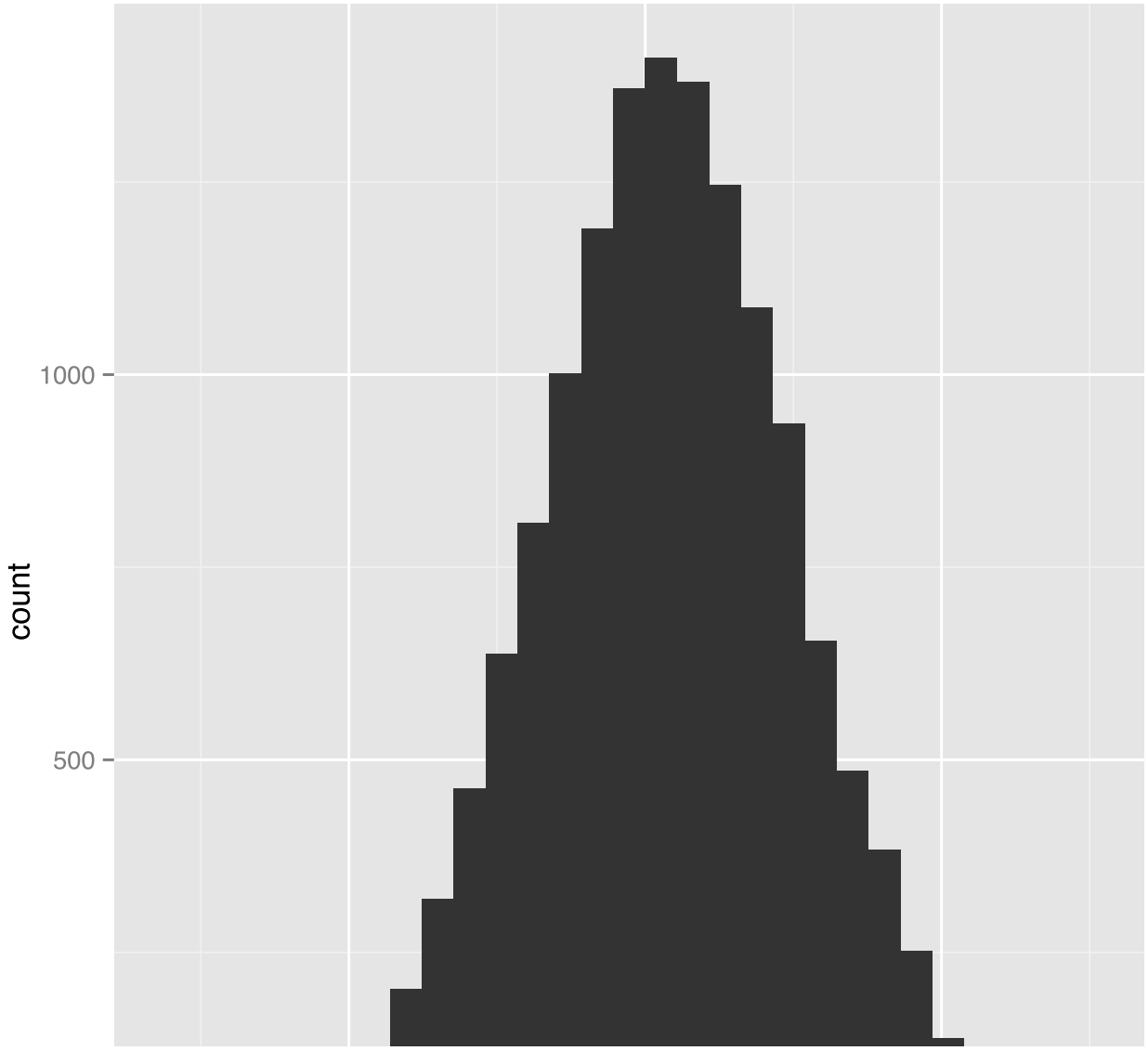


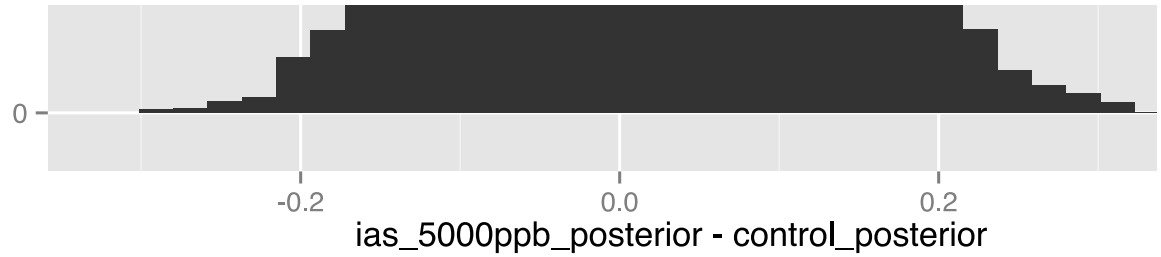
Based on this analysis, the average tumor incidence in the 5000ppb exposed group was 29%. This is similar to the result seen in Waalkes et al. Next we performed the difference analysis.

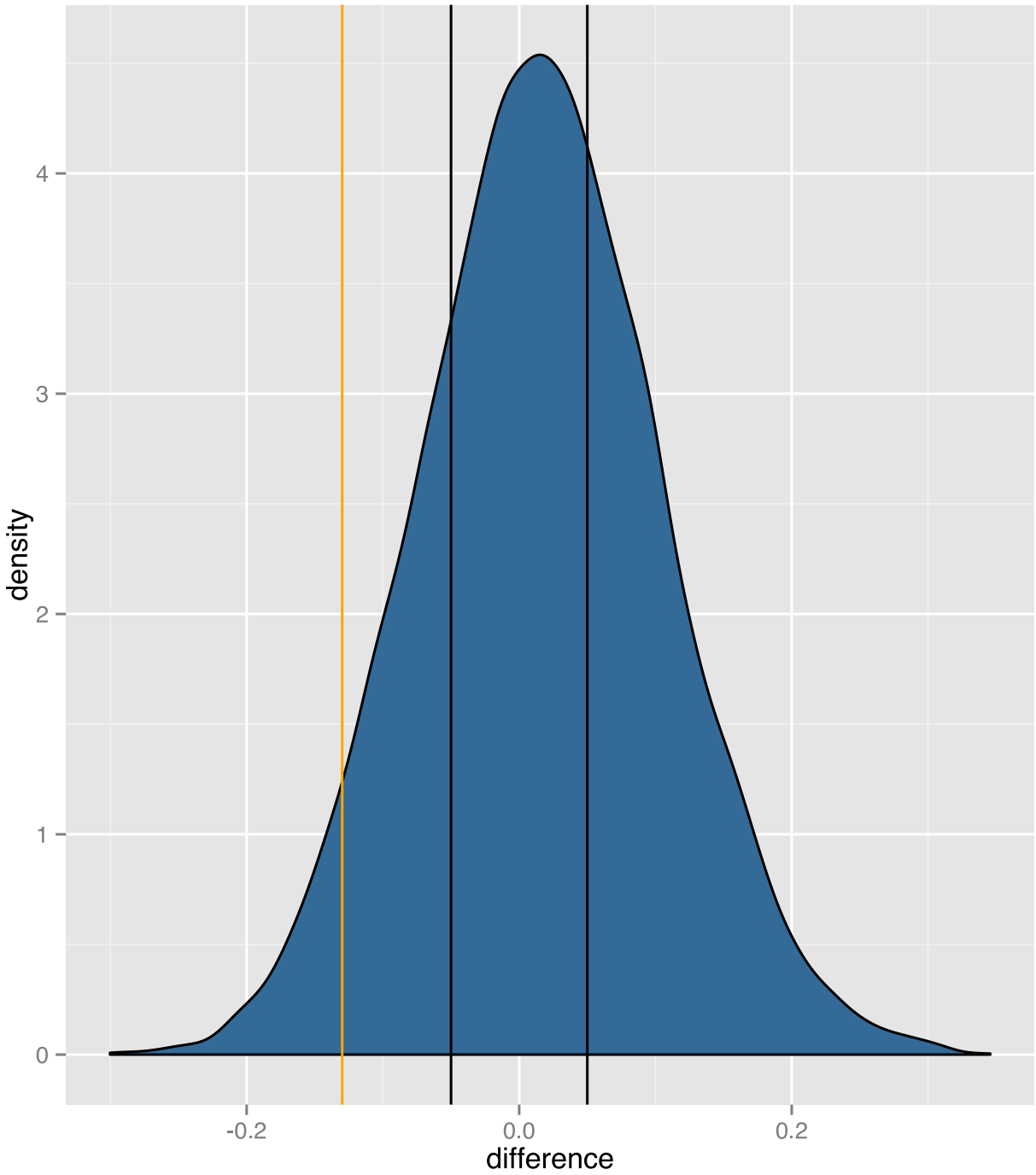
In [15]:
```
control_posterior <- extract(control_tumor_stanFit)[[1]]
ias_5000ppb_posterior <- extract(stanFit5000ppb)[[1]]
qplot(ias_5000ppb_posterior - control_posterior, geom="histogram")
diff_distro <- as.data.frame(ias_5000ppb_posterior - control_posterior
colnames(diff_distro) <- "difference"

#How much of the distribution is within the ROPE?
diff_ecdf <- ecdf(diff_distro[,1])
diff_ecdf(.05) - diff_ecdf(-.05)
#0.43%

ggplot(diff_distro, aes(difference, fill=1)) +
  geom_density(alpha = 1) +
  geom_vline(xintercept = -0.05) +
  geom_vline(xintercept = 0.05) +
  geom_vline(xintercept = quantile(diff_distro[,1], probs=c(0.05)), co
```

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adju
st this.
```

Out[15]: 0.425833333333333







The 95% HDI is inside of the ROPE, thus the null hypothesis is accepted. Based on our decision criteria, we find that 5000ppb inorganic arsenic exposure does not cause an increase in the lung adenoma and carcinoma tumor incidence in mice.

In [16]:
```
#odds ratio
odds_exposed_tumor <- .29/(1-.29)
odds_non_exposed_tumor <- .28/(1-.28)
odds_ratio <- odds_exposed_tumor / odds_non_exposed_tumor
odds_ratio
```

Out[16]: 1.05030181086519

Said another way, we feel that the odds of a mouse exposed to 5000ppb inorganic arsenic developing lung tumors 1.1:1 compared to a non-exposed mouse.

Discussion

In Cohen et al (2014 and 2015), the authors argue that had Waalkes et al considered the historical control data Waalkes et al would have seen there is no treatment-related response. Specifically, Cohen et al (2014) state "[t]he peculiar, conflicting results are likely due to the high variability in incidence of lung tumors in CD-1 mice."

In our first analysis we assessed the similarity/differences between the Tokar et al and Waalkes et al control tumor incidences. Our analysis did not support the argument from Cohen et al. In fact, our analysis suggests that although there is a 13% difference in control tumor incidences, in reality, this difference is likely to be due to simple random sampling. Our Bayesian analysis reflects the fact that the data are likely being sampled from the same distribution. In that case, it is not uncommon, and in fact would be expected, that one would see this level of difference. Thus, we are satisfied that there is not sufficient evidence to support Cohen et al's argument. However, we do concur with Cohen et al in that Tokar et al's control tumor incidence should be used in the analysis of Waalkes et al data. Thus, we used Tokar et al's control tumor incidence posterior probability that we calculated as the prior distribution for our Bayesian analysis of Waalkes et al's data.

We had previously analytically solved that Tokar et al's data could be modeled using a Beta(10, 19) distribution. Using that as our prior, we analyzed each of the inorganic arsenic exposure levels (50, 500, 5,000ppb) and constructed posterior distributions for each concentration. Using the ROPE and HDI approach, we confirmed Waalkes et al's previous results that there is a significant increase in adenomas and carcinomas in male mice at 50 and 500ppb, but not at 5,000ppb.

Thus, we have accounted for prior knowledge of the background tumor incidence in CD1 mice in Waalkes' lab using data from Tokar et al, and still confirmed the results from Waalkes et al. In so doing, we have addressed Cohen et al's chief concern in the Waalkes analysis of not accounting for the variability in CD1 mouse tumor incidences in Waalkes' lab. Furthermore, we have confirmed that inorganic arsenic exposure is increasing the lung tumor incidence in CD1 mice at 50 and 500ppb, but not at 5,000ppb. Thus, we have addressed Cohen et al's analytical concerns, and affirmed the findings from Waalkes et al.

Conclusion

Based on our analyses, we have rejected the argument from Cohen et al that the Waalkes et al study did not reproduce the control tumor incidence from the Tokar et al study. Furthermore, using Tokar et al as prior knowledge, we were able to affirm the results from Waalkes et al. Specifically, that 50 and 500ppb inorganic arsenic exposure increased lung tumor incidences in males by 2.7:1 and 3.0:1 odds, respectively. In addition, we affirmed Waalkes et al's analysis that 5,000ppb inorganic arsenic exposure did not increase lung tumor incidences in male CD1 mice, with odds of only 1.1:1. Thus, assuming there are no issues with the experimental data, we find that Waalkes et al's study demonstrates low dose inorganic arsenic exposures increases in lung tumor incidences in CD1 male mice.

Acknowledgements

The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the U.S. Army or the U.S. EPA.

References

1. Cohen SM, Arnold LL, Klaunig JE, et al. Letter to the Editor Arch Toxicol. 2014 doi: 10.1007/s00204-014-1328-1.
2. Cohen SM, Arnold LL, Klaunig JE, et al. Letter to the Editor Arch Toxicol. 2015 doi: 10.1007/s00204-015-1614-6.
3. Tokar, et al. Carcinogenic effects of “whole-life” exposure to inorganic arsenic in CD1 mice. Toxicol Sci. 2011;119:73–83.
4. Waalkes MP, Qu W, Tokar EJ, et al. Lung tumors in mice induced by “whole-life” inorganic arsenic exposure at human-relevant doses. Arch Toxicol. 2014;88:1619–1629.