UPDATE 11.5.2017

Analysis of short bouts

I analyzed the bouts of length 1,2 and 3.
From my analysis it looks like there is some information there, so might not be ok to discard that data. But I will have to discuss the statistics tomorrow with Anna from Multipark.

I had to do some simplifications, to make things easier for now. I examined each of 6 conditions separately and within each condition I did not make dosage a factor inside drug, but rather took 12 drugs*3 dosages as 36 drugs, to avoid an extra level.
I then analyzed all 36 “drugs” relative to the control.

First I just plotted the mean over all subjects per time frame for each “drug” versus control, for several variables:

-total bout count for reference,
-permille or length 1,2 and 3 bouts, to get a picture of short bouts, separate from the total increase/decrease in bout count
-mean proportion of each of 8 turns in bout length 1,2 and 3

I could already observe that there are differences from the control, this was especially evident when keeping in mind the dosage factor.

Plotting is not enough tough, since it is averaged over subjects and turn proportions are not adjusted for the total count of bouts of that length, so I also had to do some sort of analysis.
I took all the variables as count data and tried with regression fitting with a negative binomial or a quasipossion distribution.
I am uncertain about some things with this analysis and I will discuss it with Anna tomorrow, but the general idea with these two types of models was that I can model count data, which is not normally distributed, has unequal variance in time and I can adjust for covariates, having control as a reference for the regression coefficients with subject being random.
Examples:

total bout count

```
#negative binomial
model_nb_theta<-summary(glm.nb(AllBoutCount~TimeFactor+Group+TimeFactor*Group))
[[18]]#estimate theta

model_nb<-glmer(AllBoutCount~TimeFactor+Group+TimeFactor*Group+(1|Subject),
family = negative.binomial(model_nb_theta))

#quasipoisson
model_quasipoisson<-glmmPQL(AllBoutCount~TimeFactor+Group+TimeFactor*Group,
random=~1|Subject,family=quasipoisson,data=all_dataset)
```

permille of length1 bouts

```
#negative binomial
model_nb_theta<-
summary(glm.nb(Length1BoutCountPerMille~TimeFactor+Group+TimeFactor*Group))
[[18]]#estimate theta

model_nb<-glmer(Length1BoutCountPerMille~TimeFactor+Group+TimeFactor*Group+(1|
Subject), family = negative.binomial(model_nb_theta))
```

```
#quasipoisson
model_quasipoisson<-
glmmPQL(Length1BoutCountPerMille~TimeFactor+Group+TimeFactor*Group, random=~1|
Subject,family=quasipoisson,data=all_dataset)
```

proportion(rounded up permille) of Cbends

```
#quasipoisson
model_quasipoisson<-
glmmPQL(round(Length1CBends*1000)~TimeFactor+Group+TimeFactor*Group+Length1BoutC
ount, random=~1|Subject,family=quasipoisson,data=all_dataset)
```

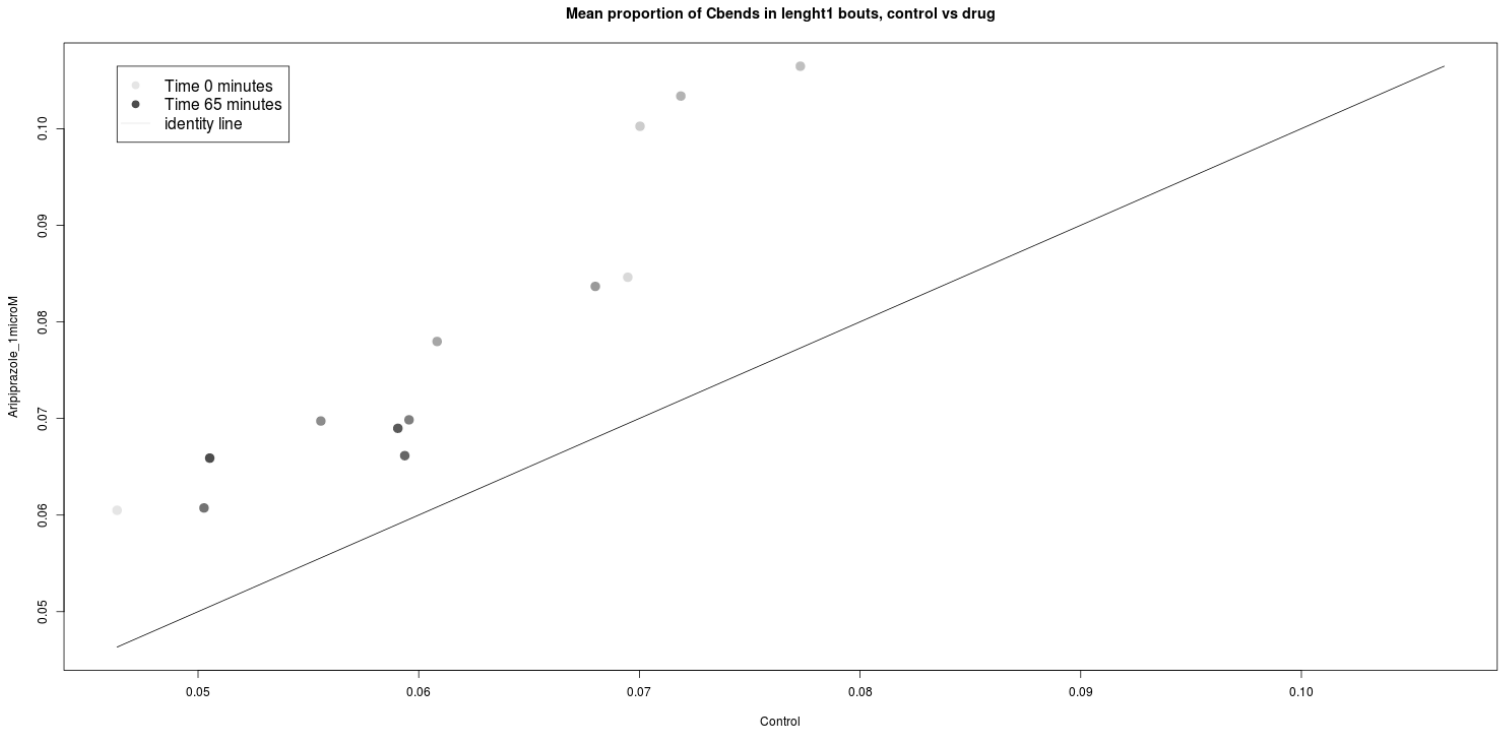
With these analyses I wanted to see if there is an overall change between the control and the drugs, change in time or both, something like a split plot in a mixed anova, where time is the within subject factor and group is the between.
There could be a difference between the groups, but the change in time would still be the same, or the groups could be the same and there is only change in time. I was expecting that if looking within the condition, I will get no difference for the overall comparison between groups, but only for the difference in change in time. But they are of course mixed and I have to check with Anna if the model and interpretation of results is ok.

For now I will not add any numbers, since they would probably mean nothing to everybody, I will just add some plots as examples and illustrations of results:

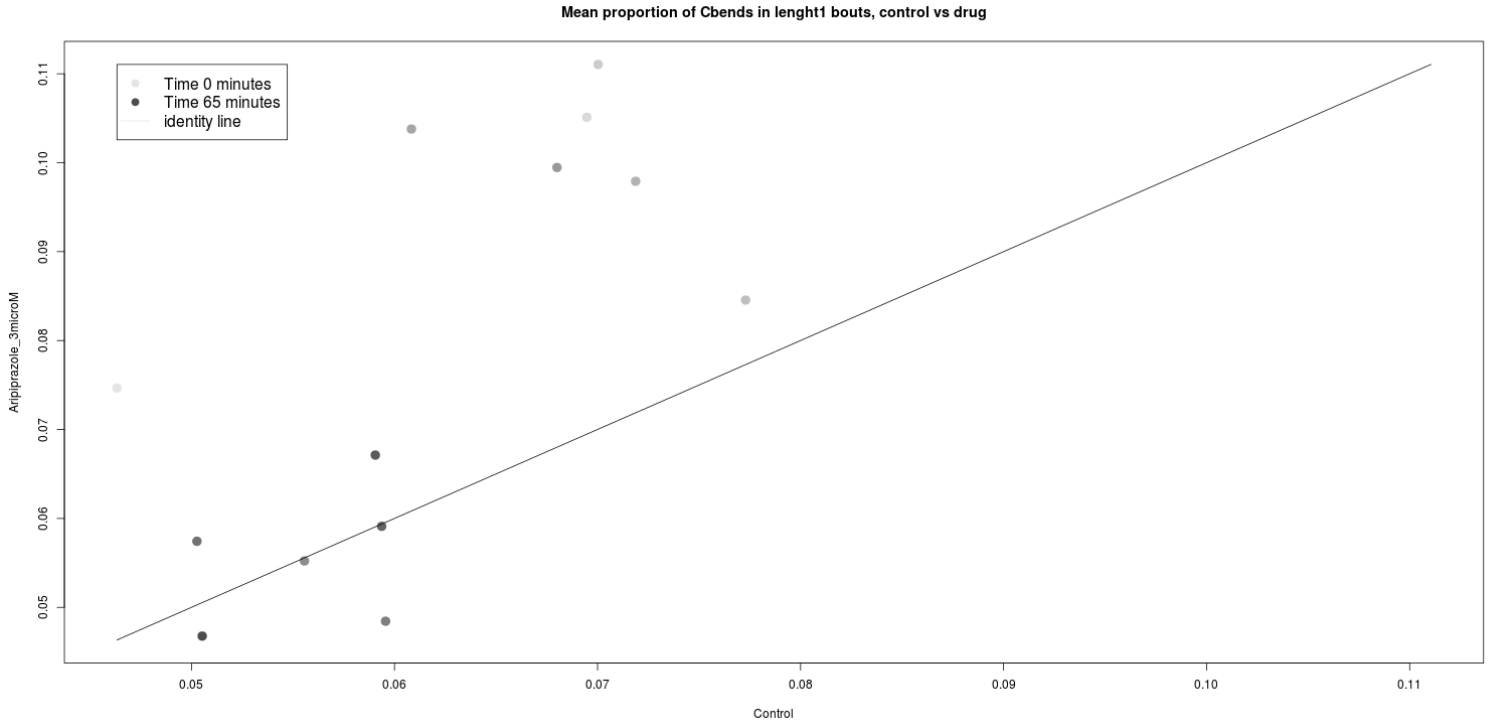
First, the mean of certain variables in time, plotting control vs drug:

Looking at proportion of Cbends for length1 bouts of Control and Aripiprazole in Dark and DarkPTZ, it seems as the dosage increase has an effect:

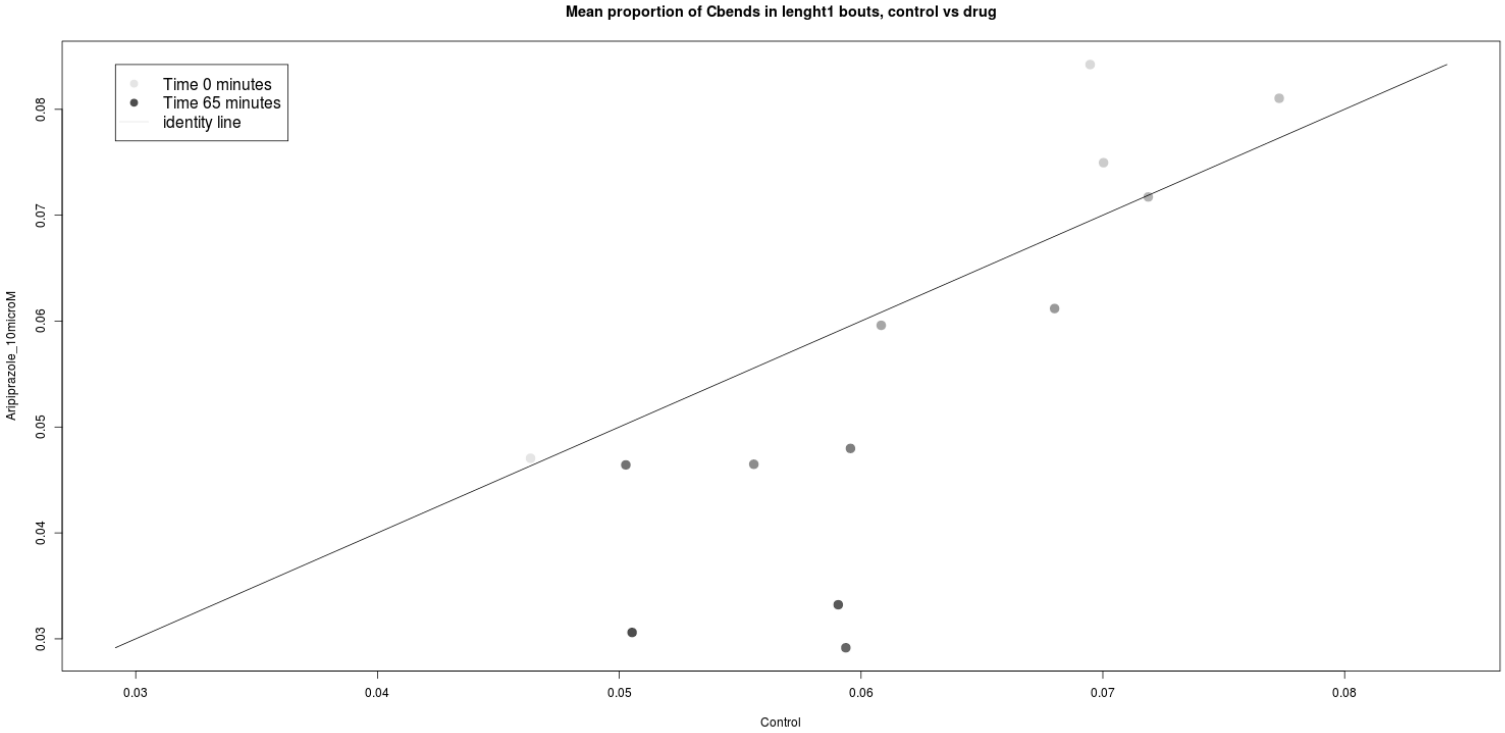
Dark, 1microM, slightly bigger proportions in Aripiprazole:



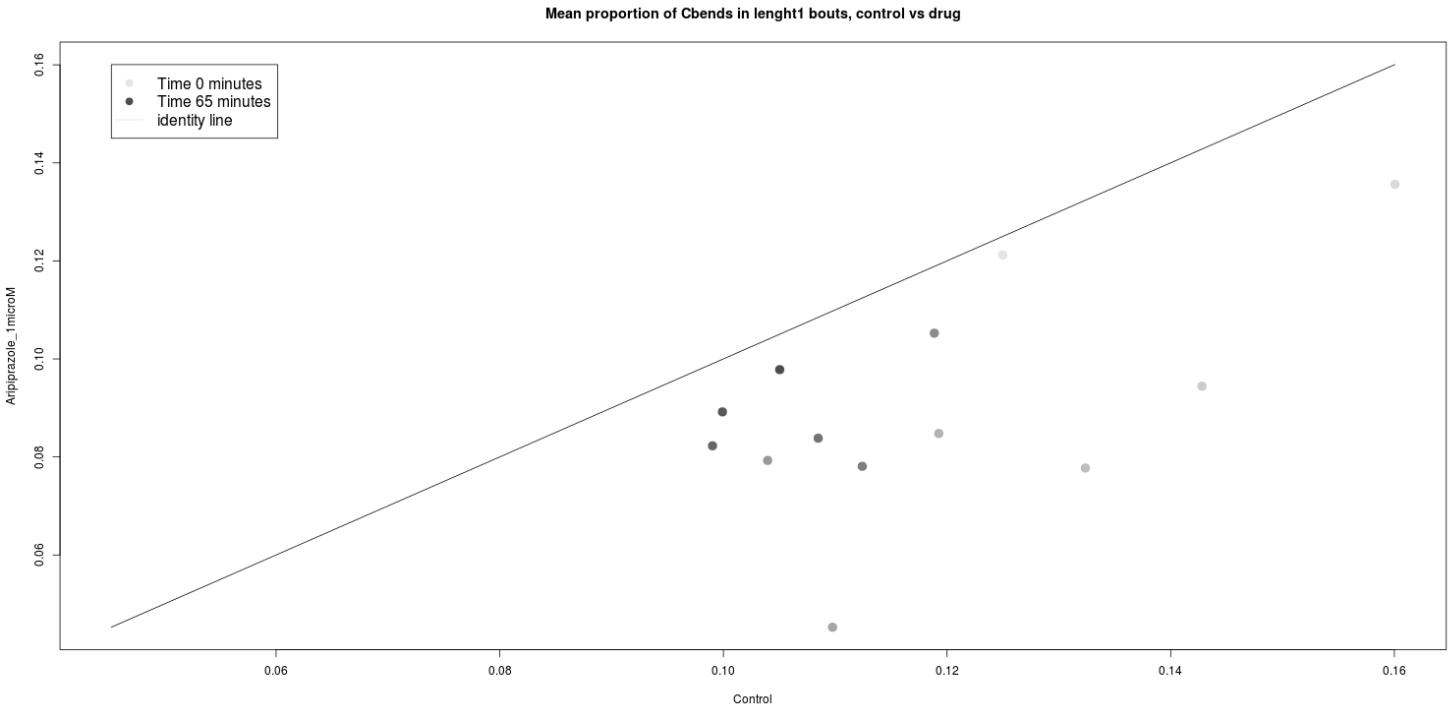
Dark, 3microM, slightly bigger proportions in Aripiprazole, but not as time passes:



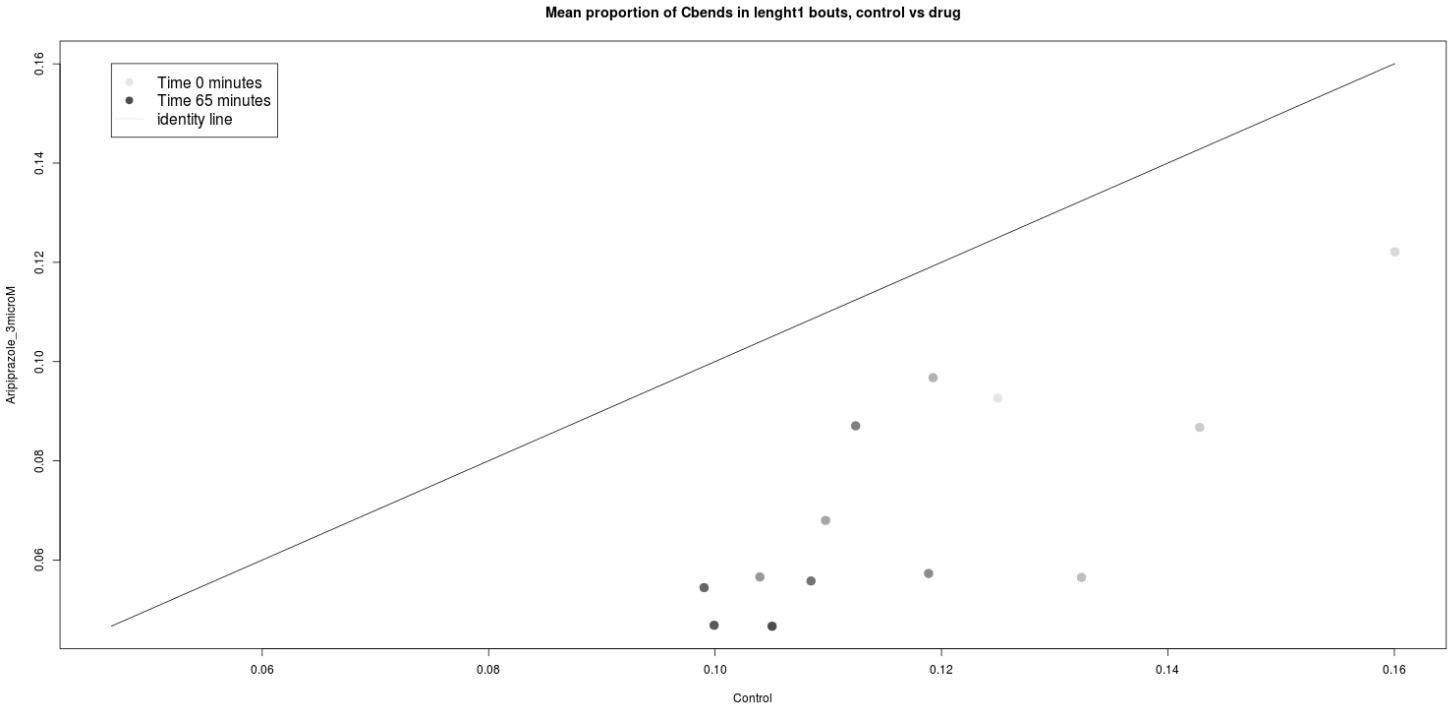
Dark, 10microM, proportions in Aripiprazole get smaller as time passes:



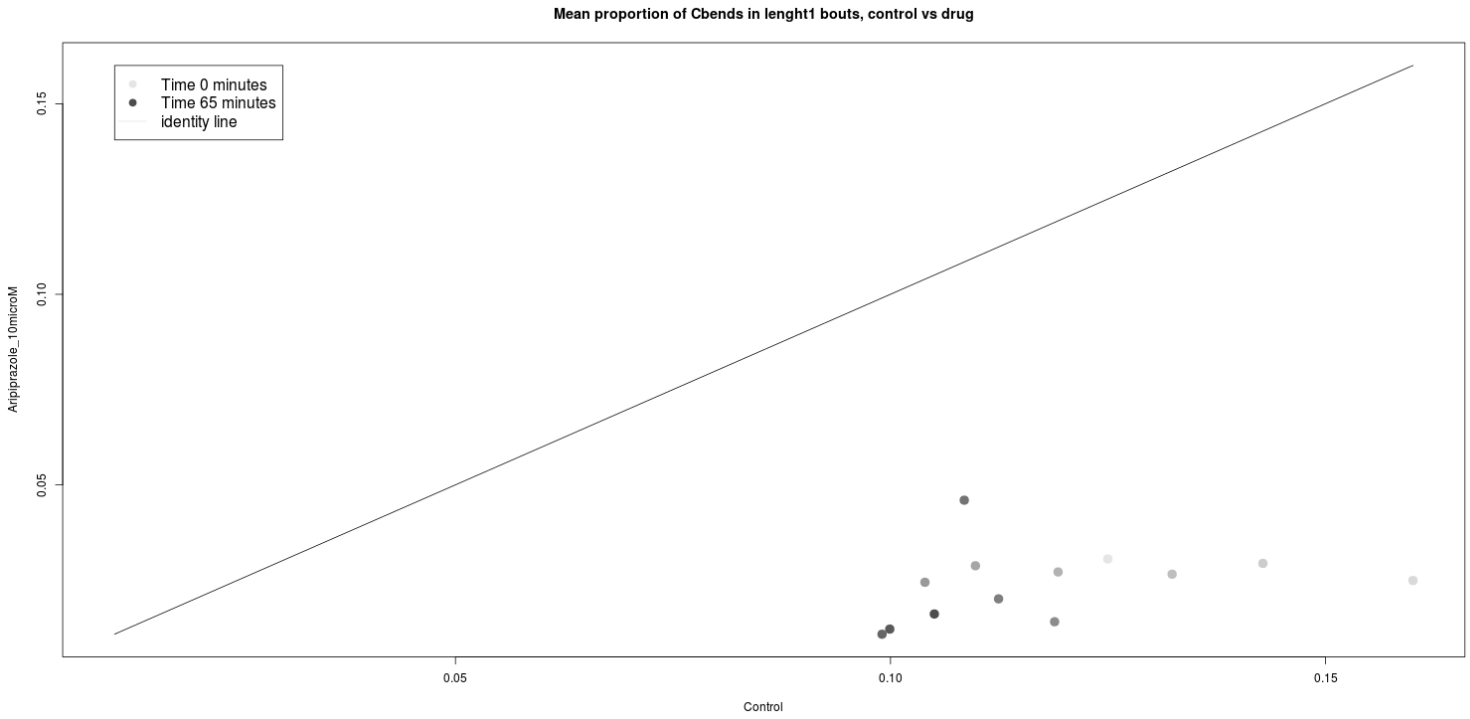
DarkPTZ, 1microM, slightly smaller proportions in Aripiprazole:



DarkPTZ, 3microM, proportions get even smaller in Aripiprazole:

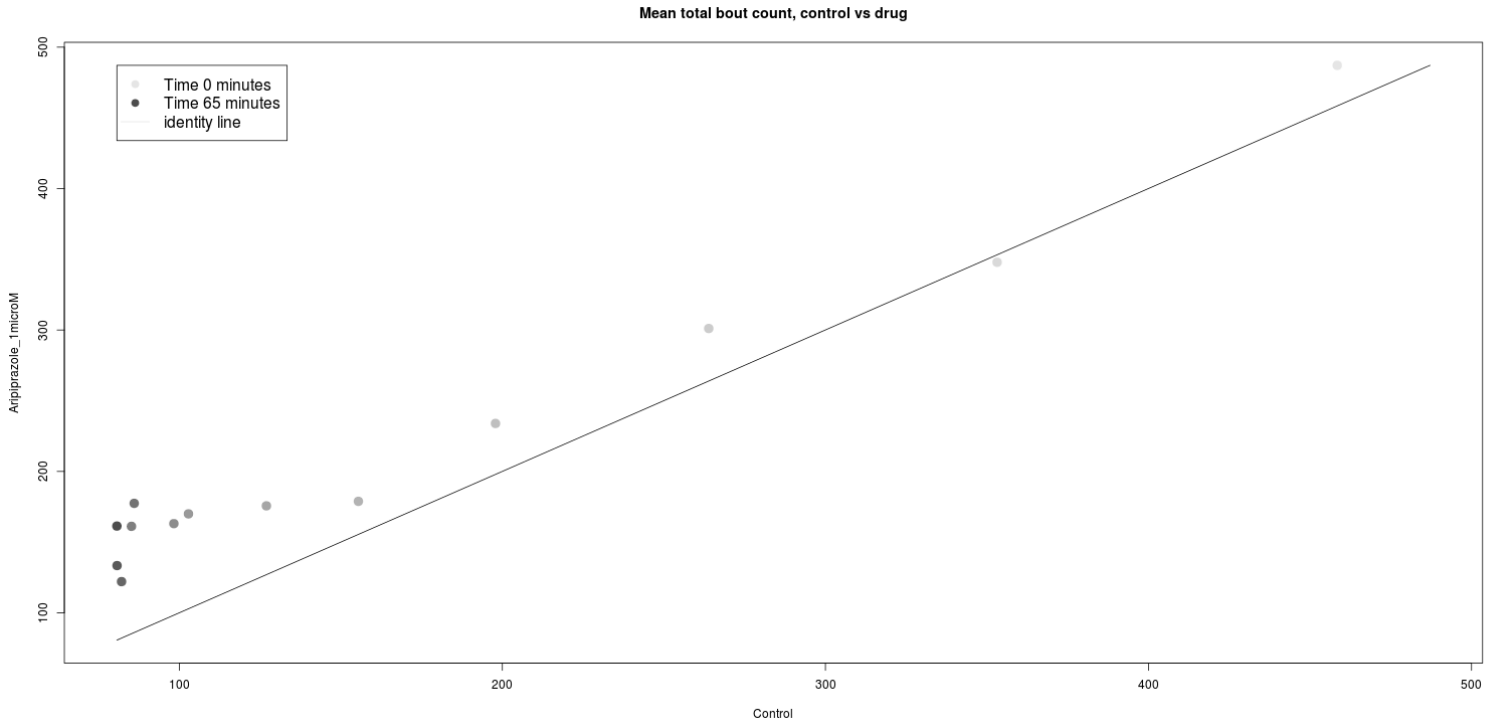


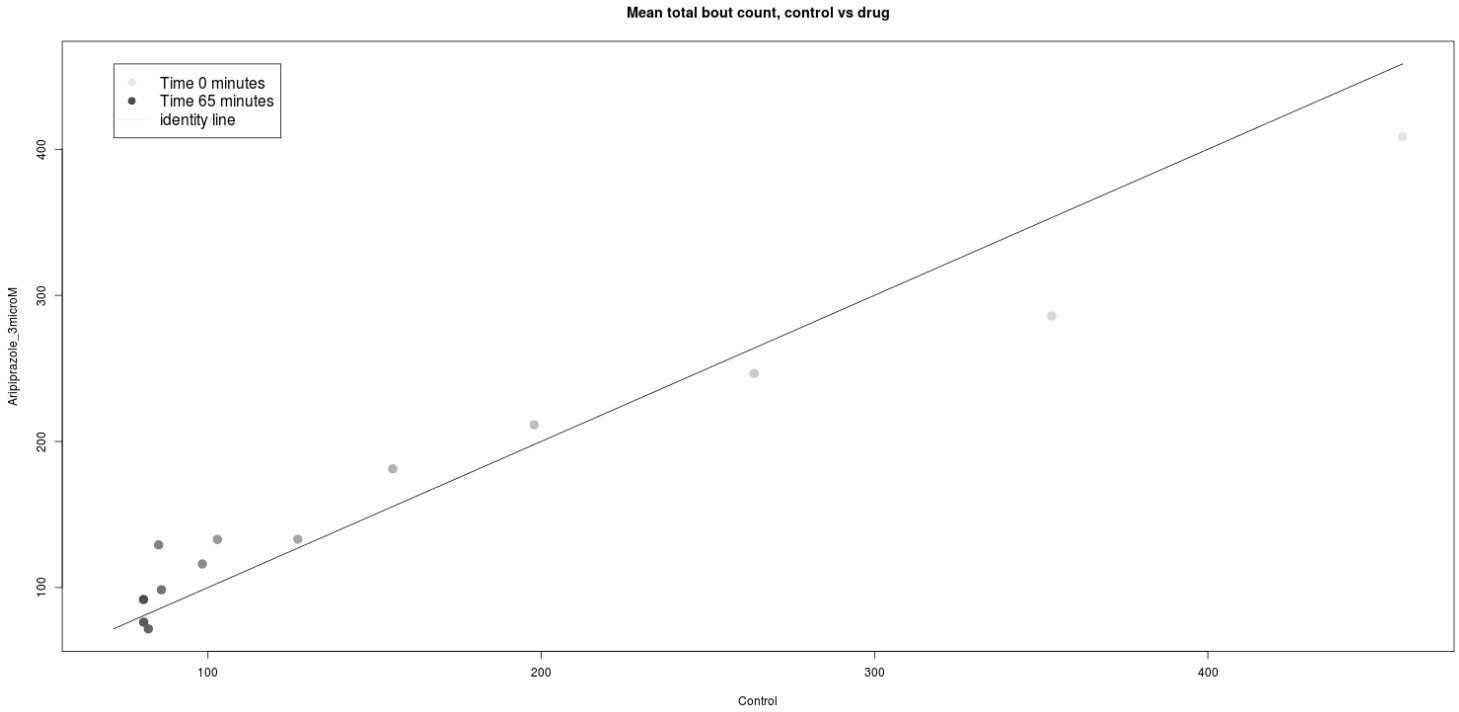
DarkPTZ, 10microM, proportions get much smaller in Aripiprazole:

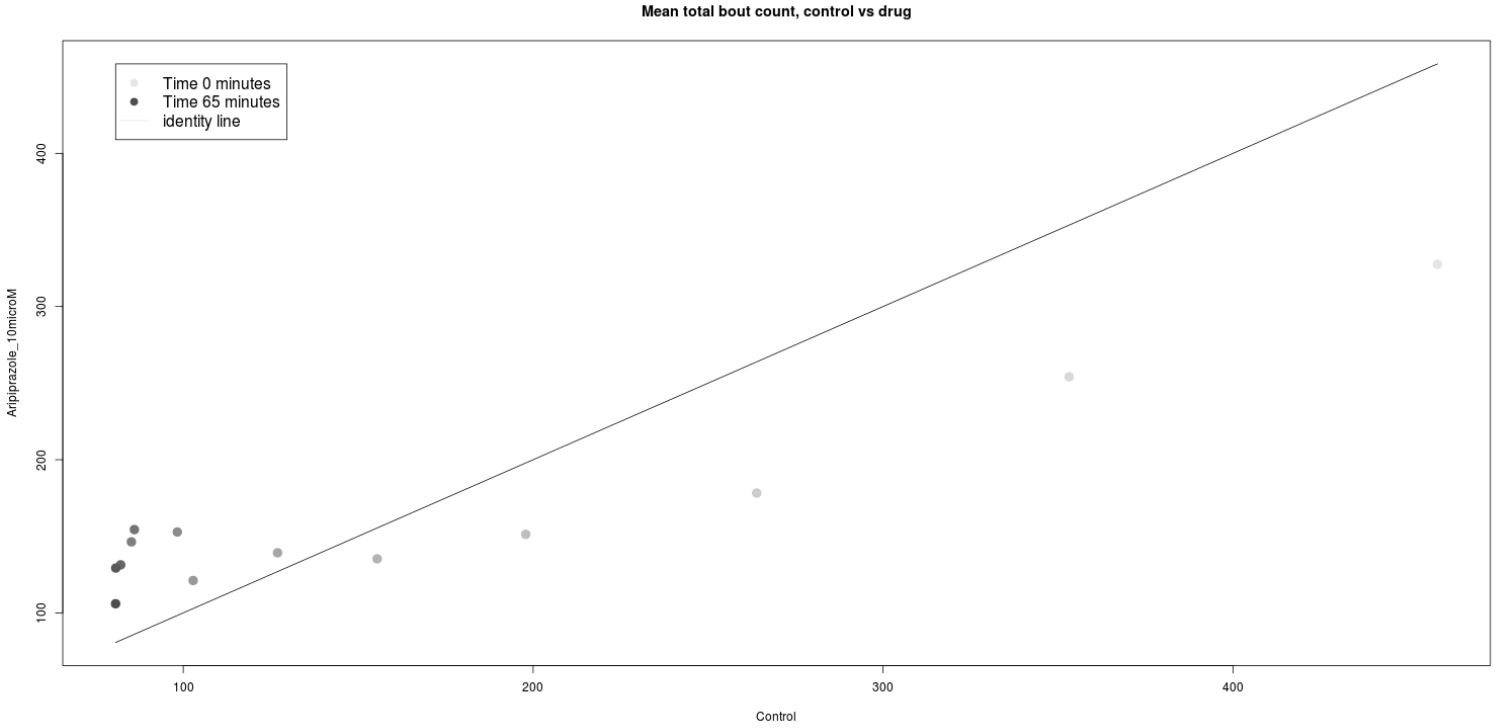


Looking at just the plots we cant be sure if these changes are significant, considering the total variation of all subjects in the proportions, since the change are for ~1-15% only. Another issues is that we can not be sure if the changes we are looking at are just the the reflection of the change in total count of that length, example for Dark and DarkPTZ, control vs Aripiprazole dosages, for total bout count and permille of length1 bouts:

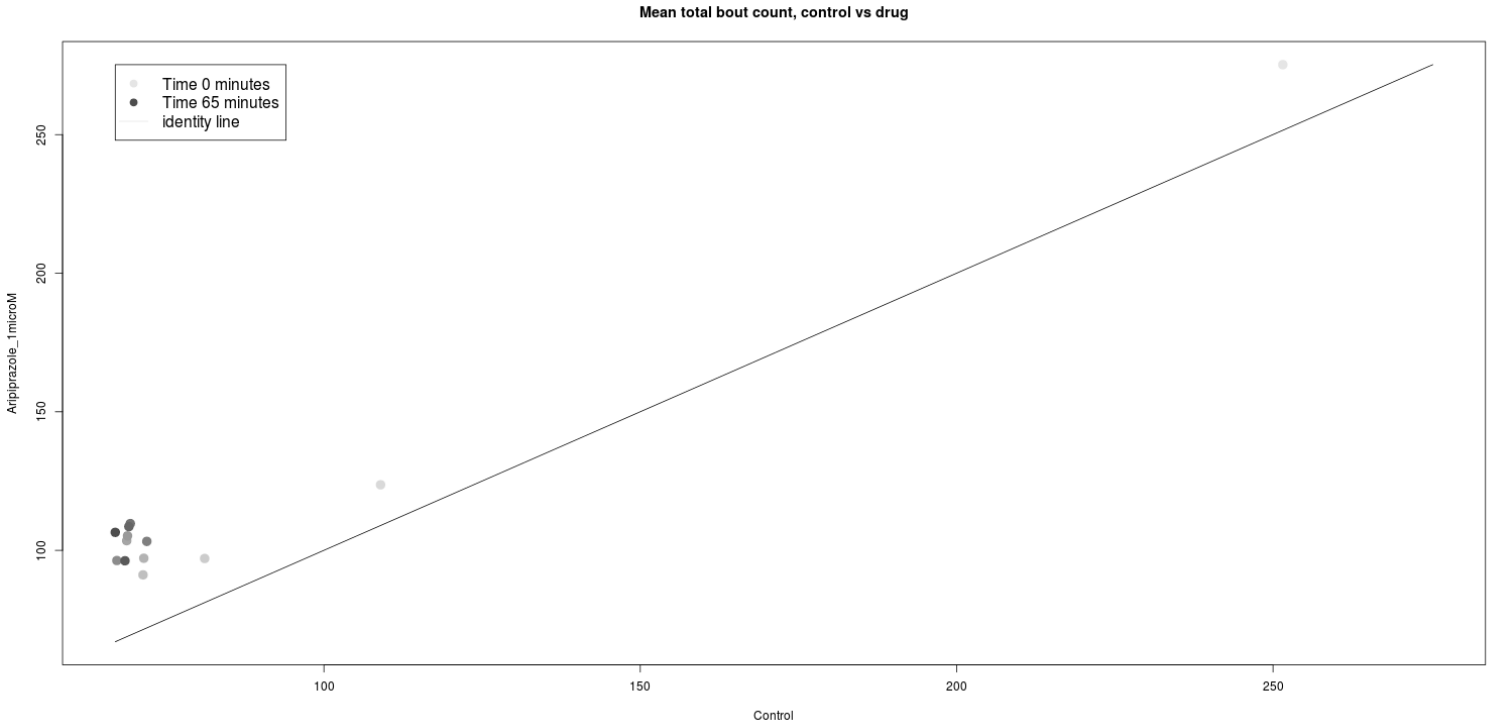
Dark, total bout count, 3 dosages:

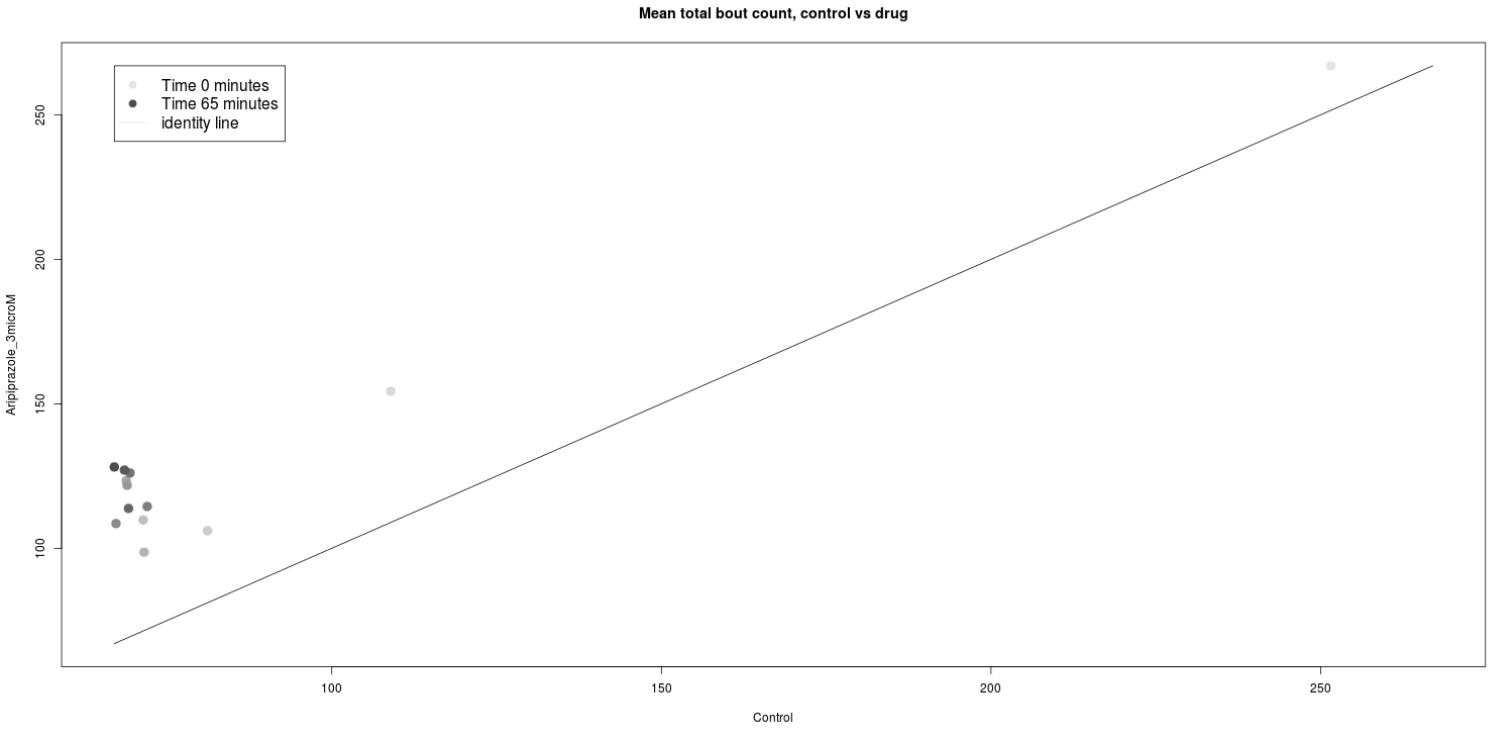


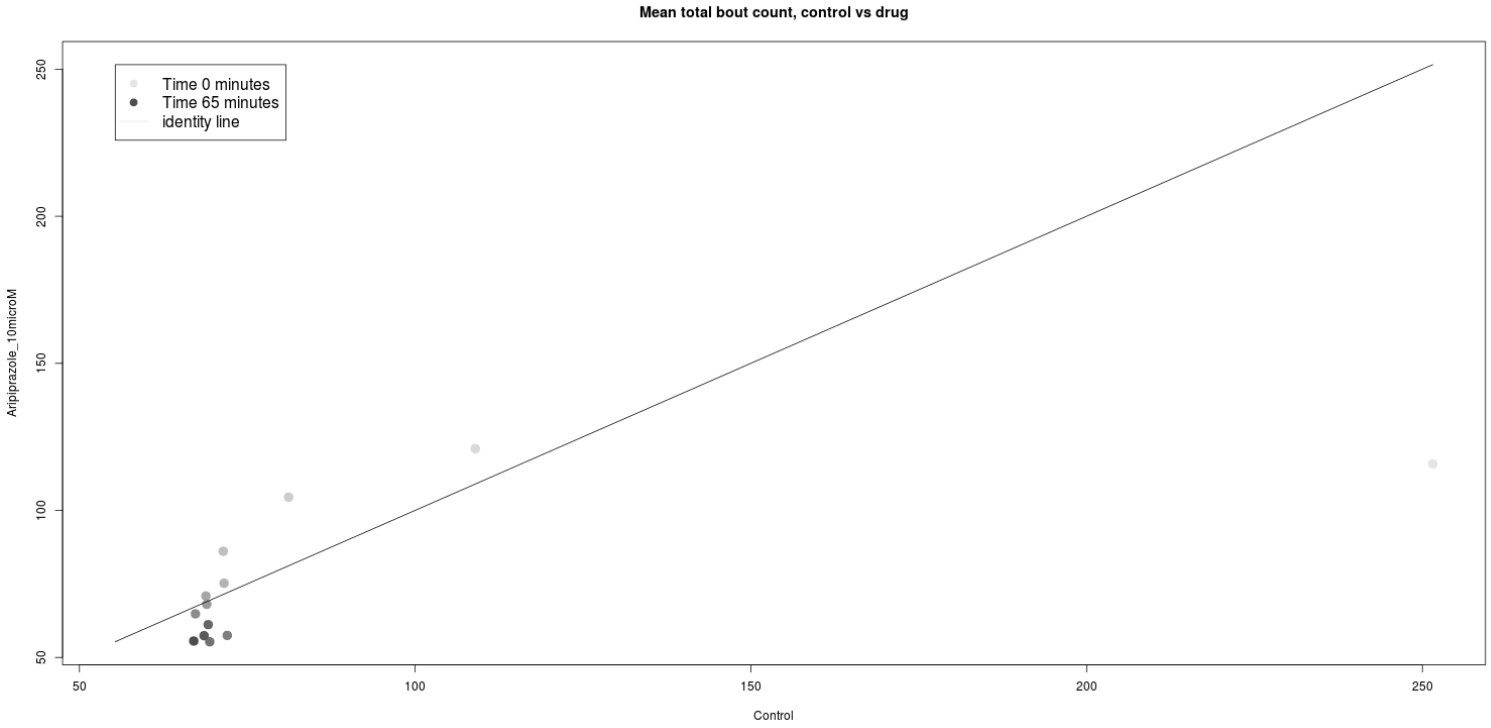

Mean total bout count, control vs drug
Time 0 minutes
Time 65 minutes
identity line
Aripiprazole_3microM
Control
100
200
300
400


Mean total bout count, control vs drug
Time 0 minutes
Time 65 minutes
identity line
Aripiprazole_10microM
Control
100
200
300
400

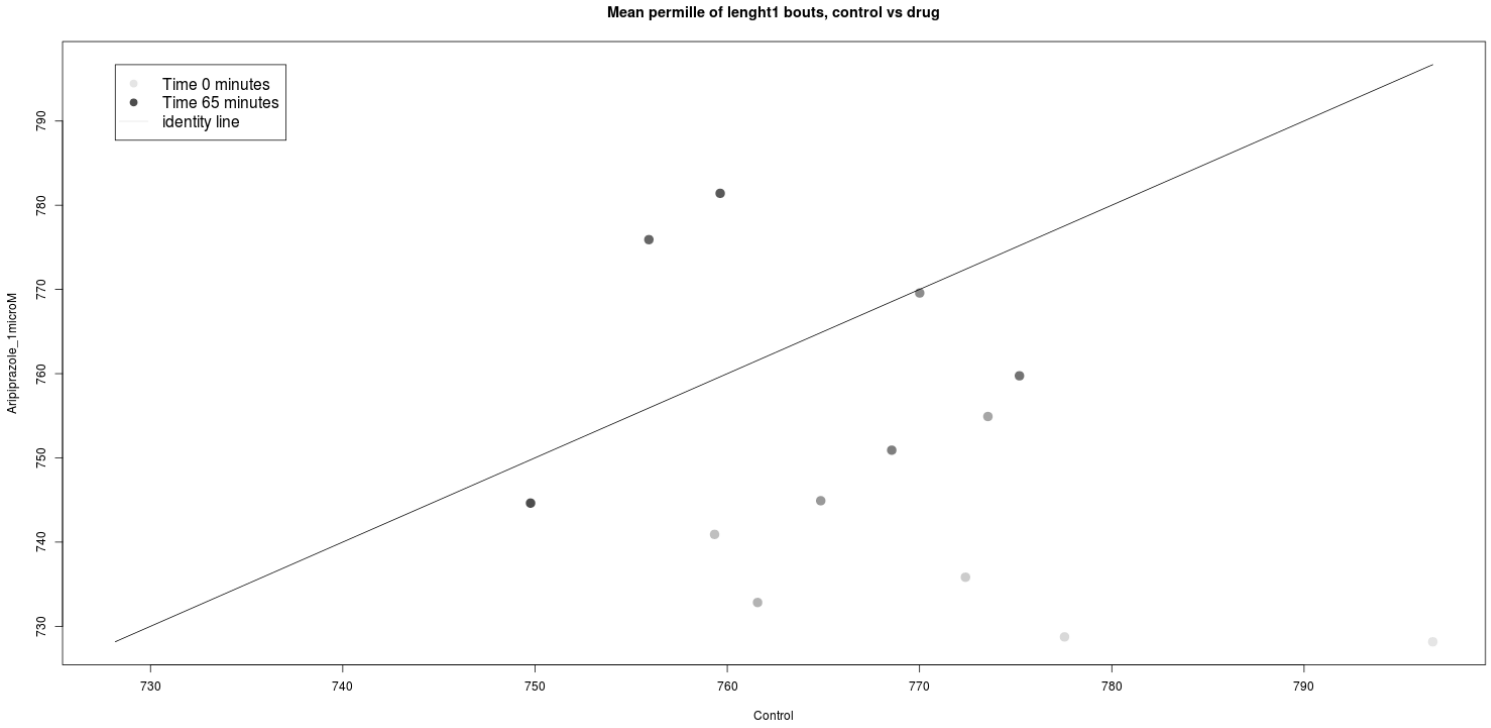
DarkPTZ, total bout count, 3 dosages:

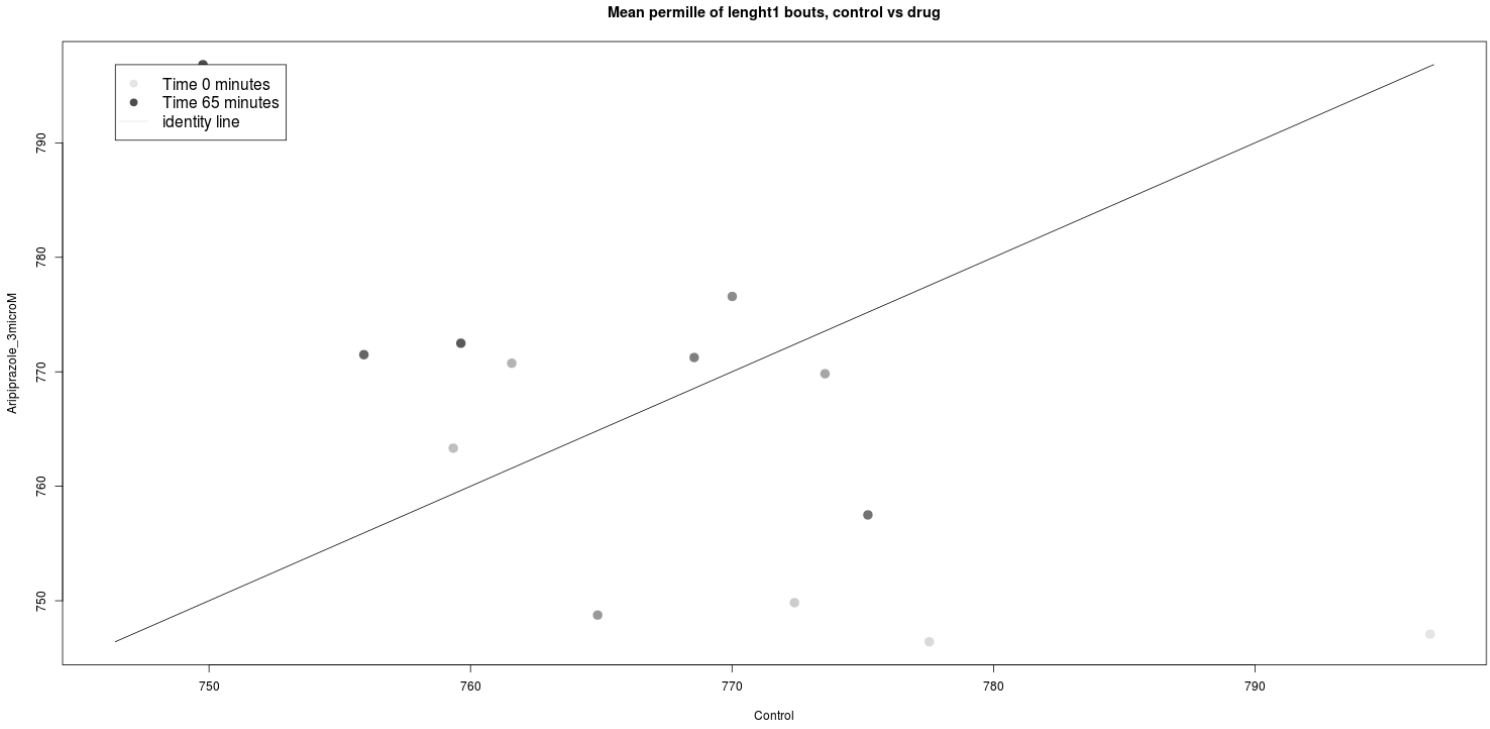


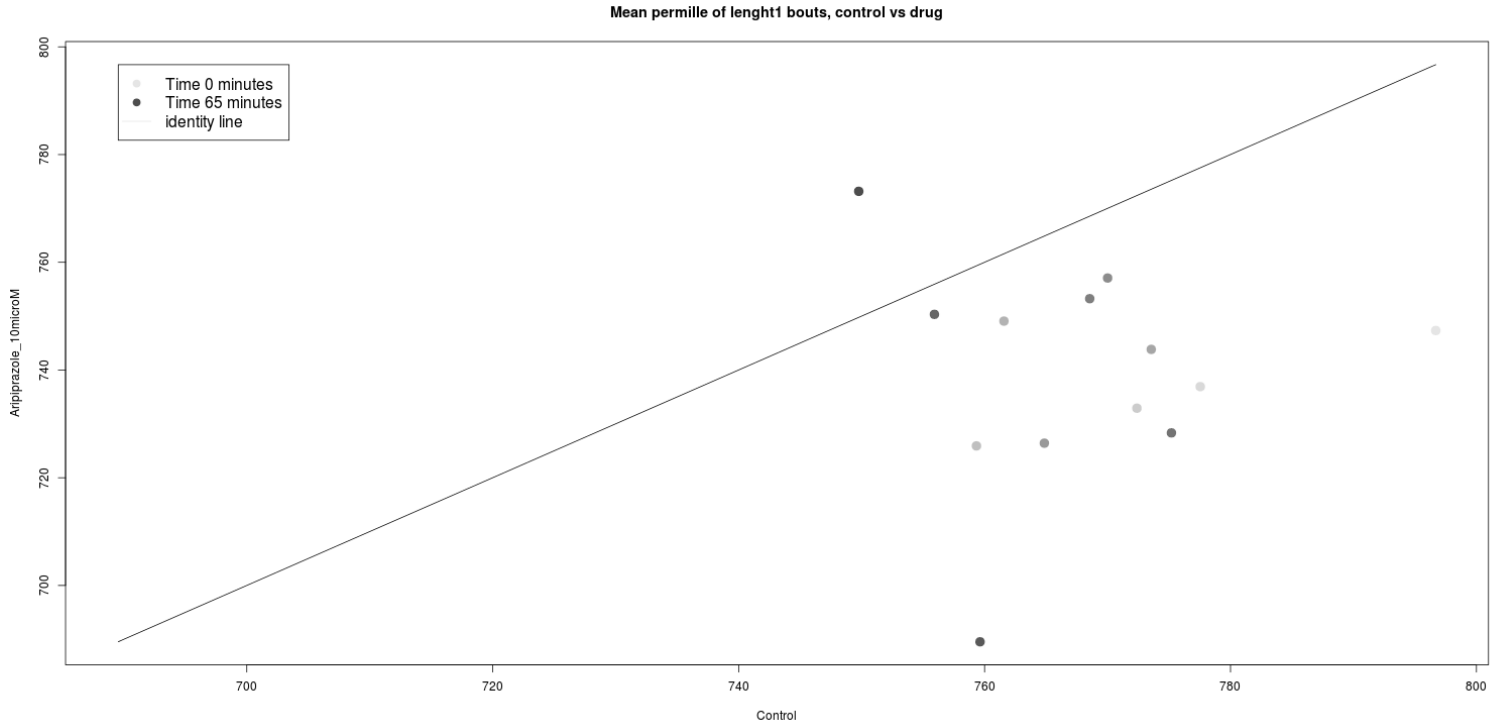




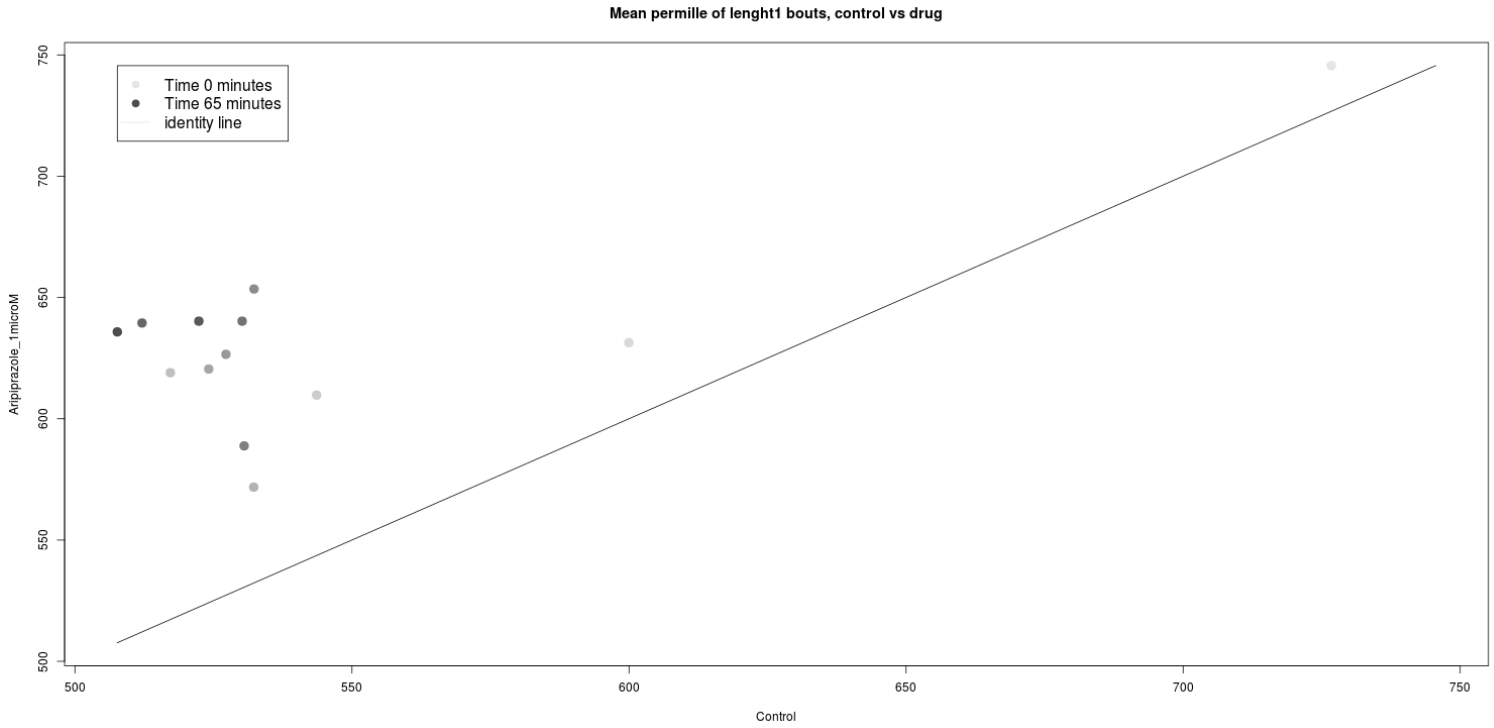
Dark , permille of length1 bouts, 3 dosages

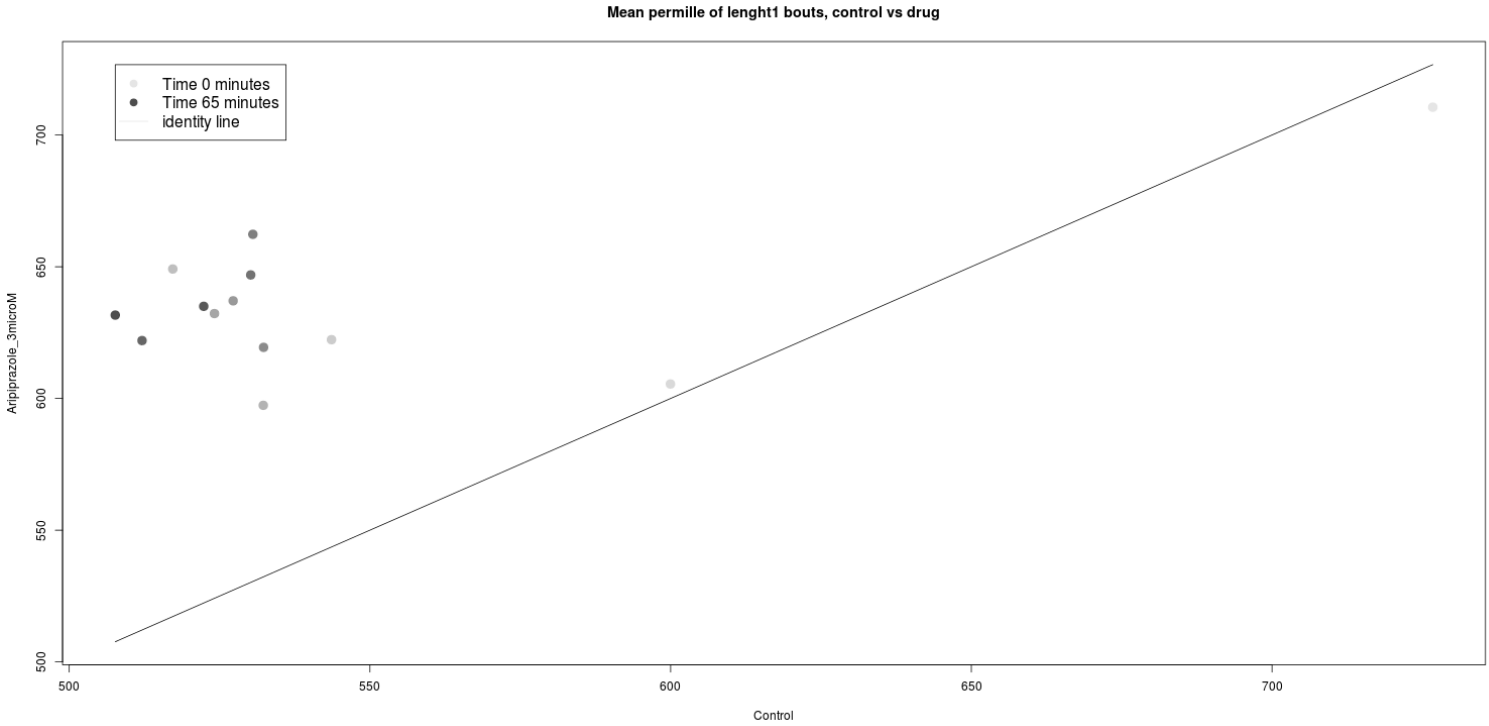


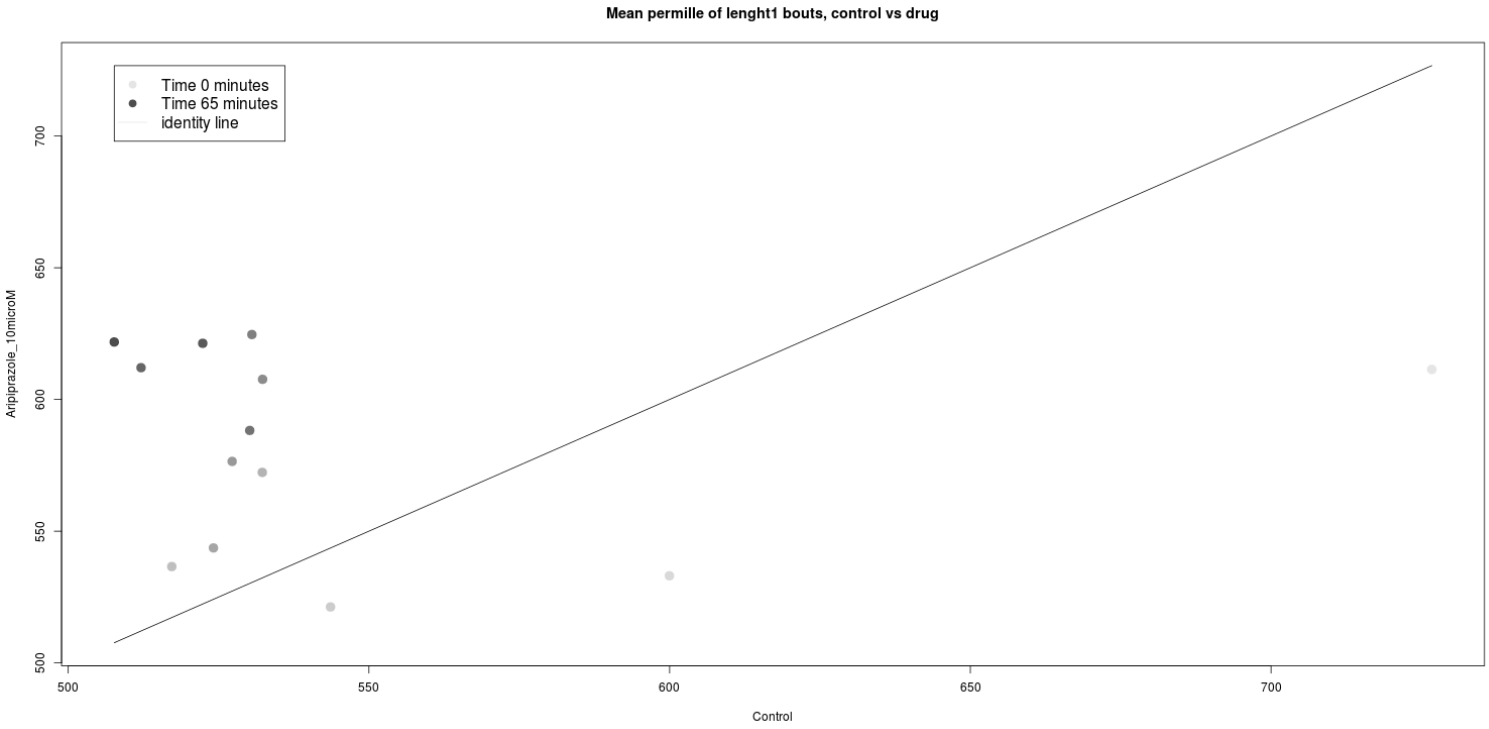

Mean permille of lenght1 bouts, control vs drug
Time 0 minutes
Time 65 minutes
identity line
Aripiprazole_3microM
Control
750
760
770
780
790


Mean permille of lenght1 bouts, control vs drug
Time 0 minutes
Time 65 minutes
identity line
Aripiprazole_10microM
Control
700
720
740
760
780
800

DarkPTZ , permille of length1 bouts, 3 dosages

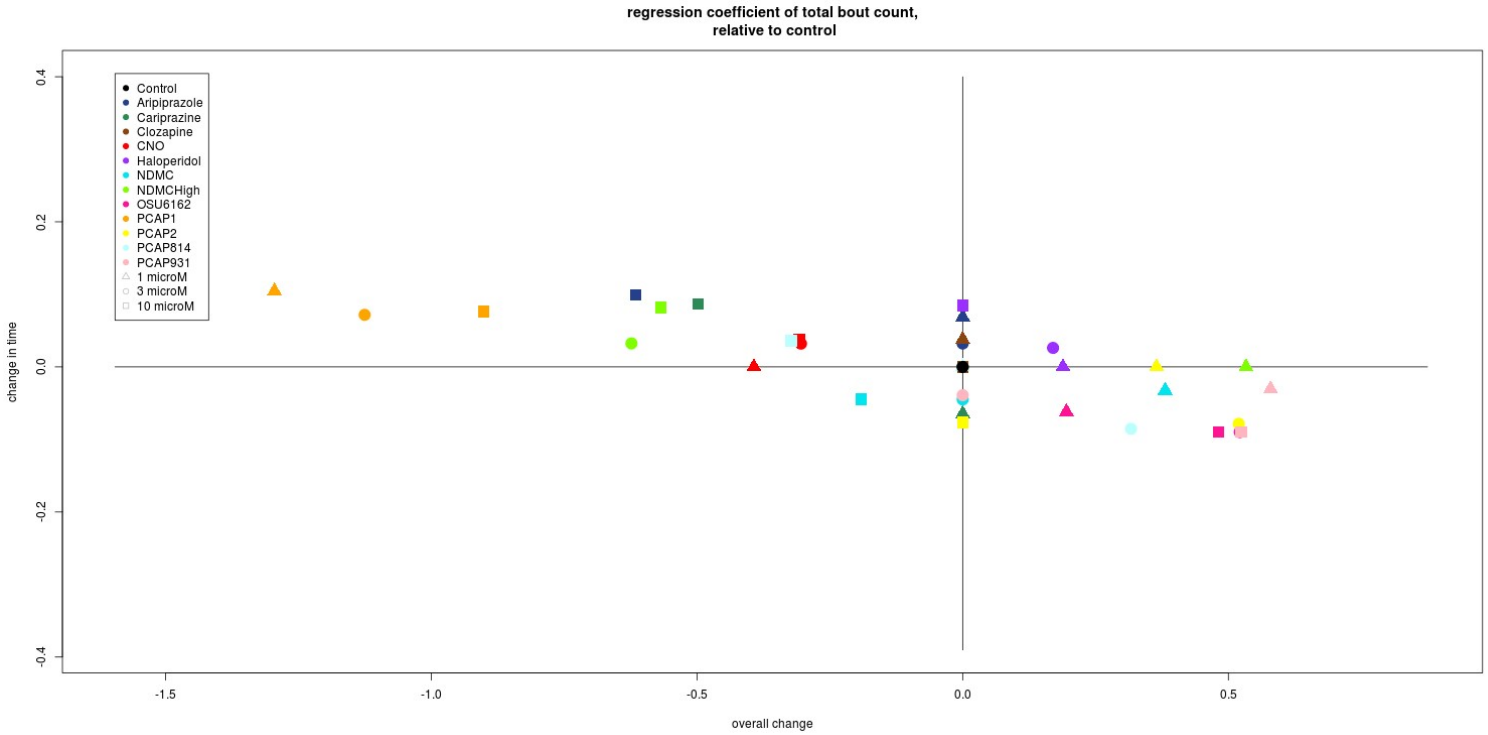




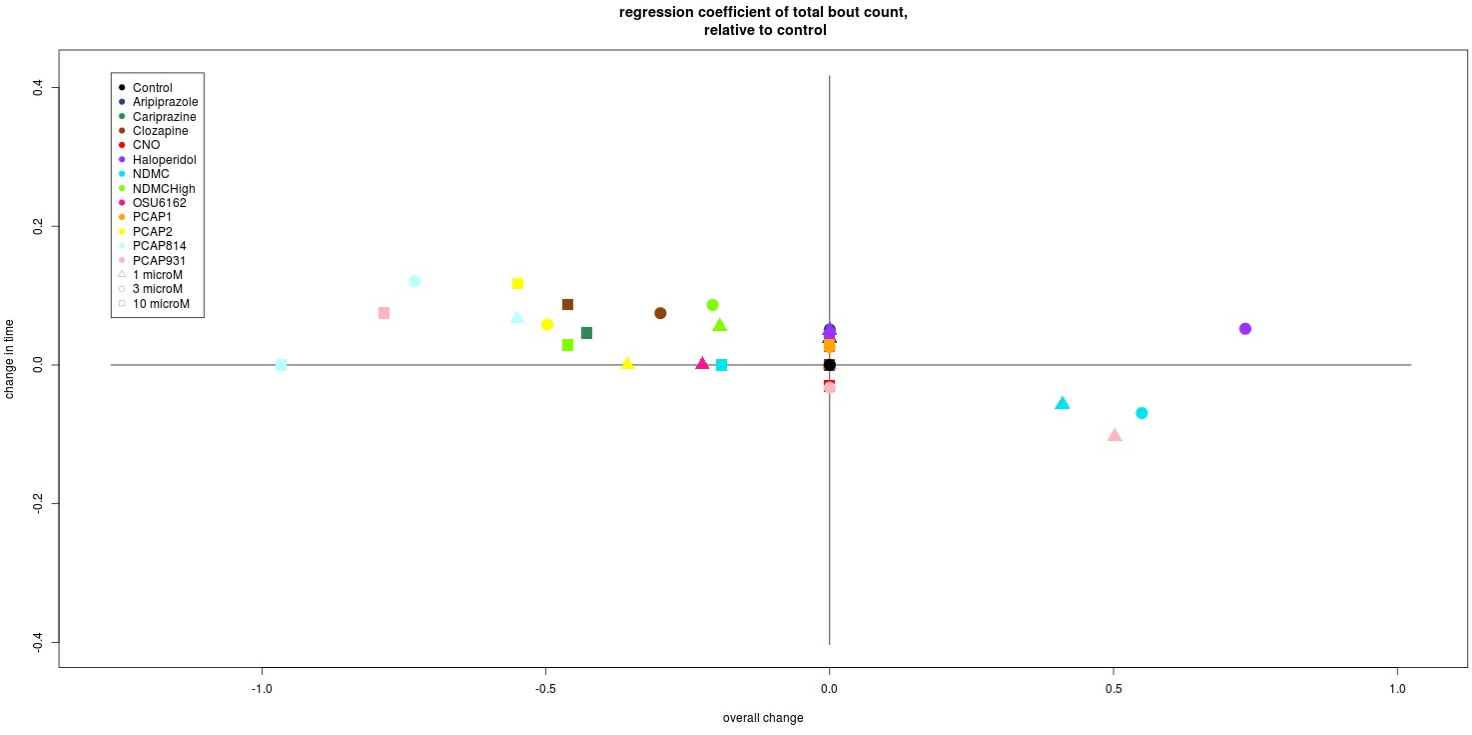

Mean permille of lenght1 bouts, control vs drug
Time 0 minutes
Time 65 minutes
identity line
Aripiprazole_10microM
Control
500
550
600
650
700

So I plotted the results from the regression, if the coefficient was non significant it was set to 0, control was 0 as default.
I plotted coefficients for different variables looking at the overall change in x and change in time in y, having different colors for drugs and symbols for dosages.
From these results, it looks like there is change from the control in some drugs, either overall change, change in time or both.
Just from looking at one drug, the no change could mean that the drug is ineffective or bouts of length one have no information, but since there is change in some variables, we could conclude that no change is due to the drug having no effect, since bouts of length 1 can have information in some drugs.

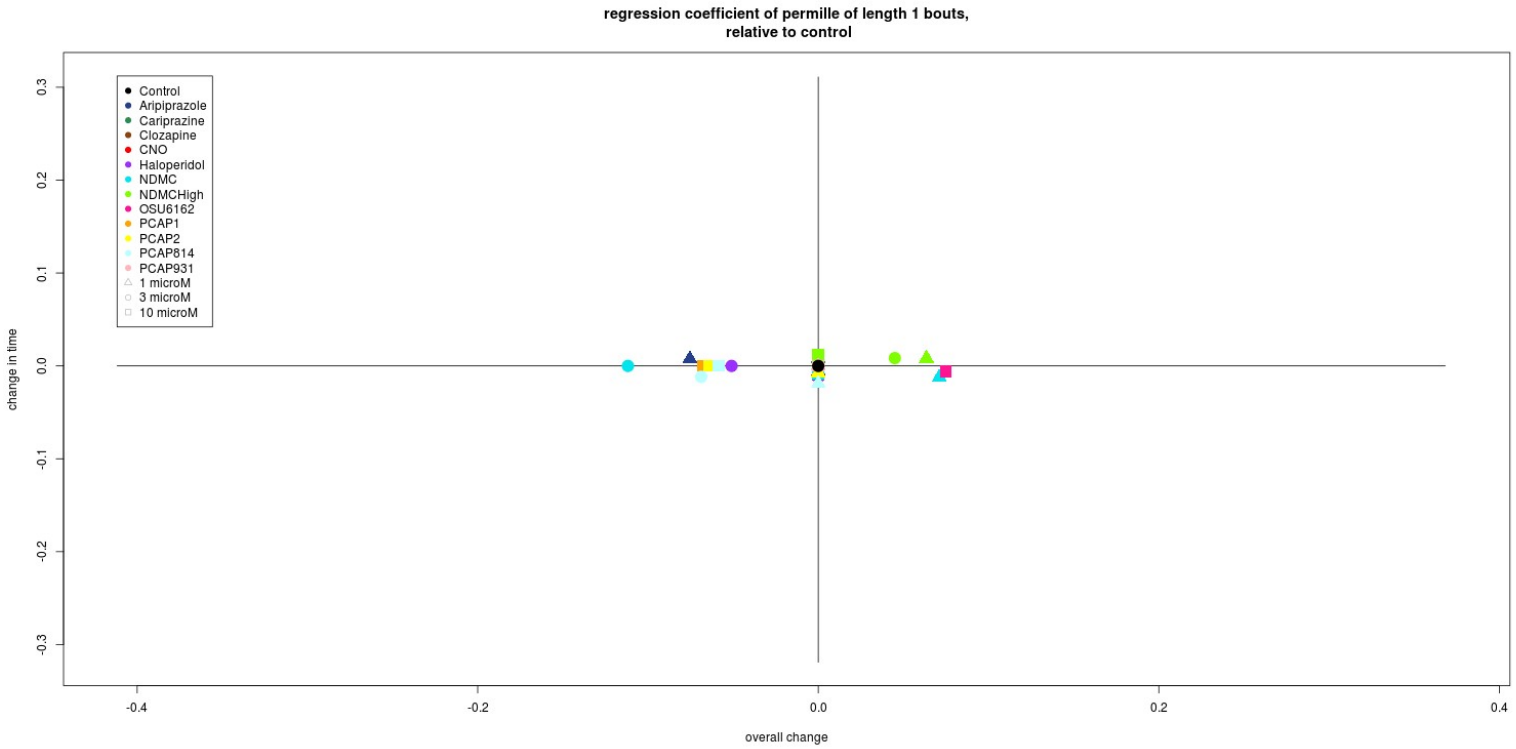
Dark, total bout count:



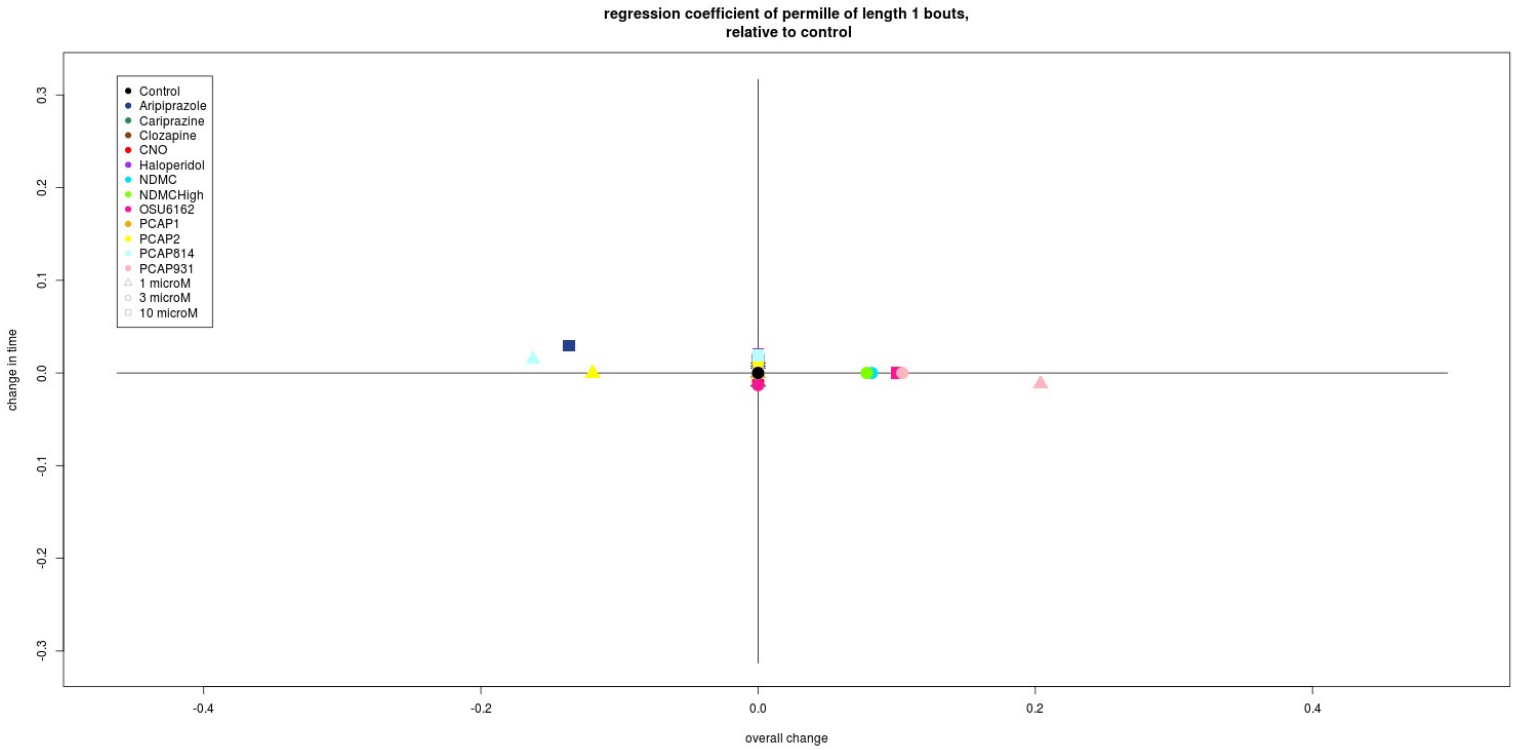
DarkPTZ, total bout count:



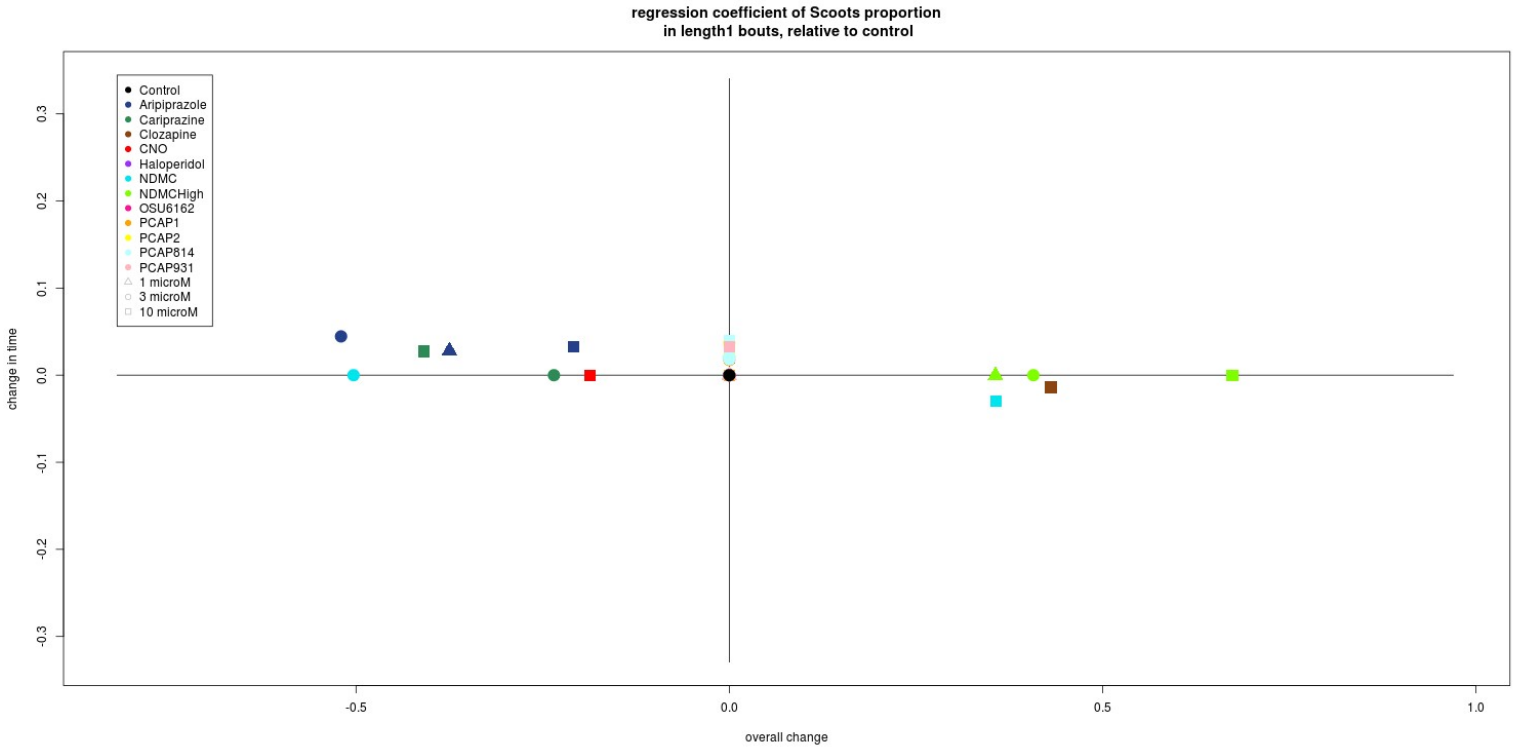
Dark, permille of length1 bouts:



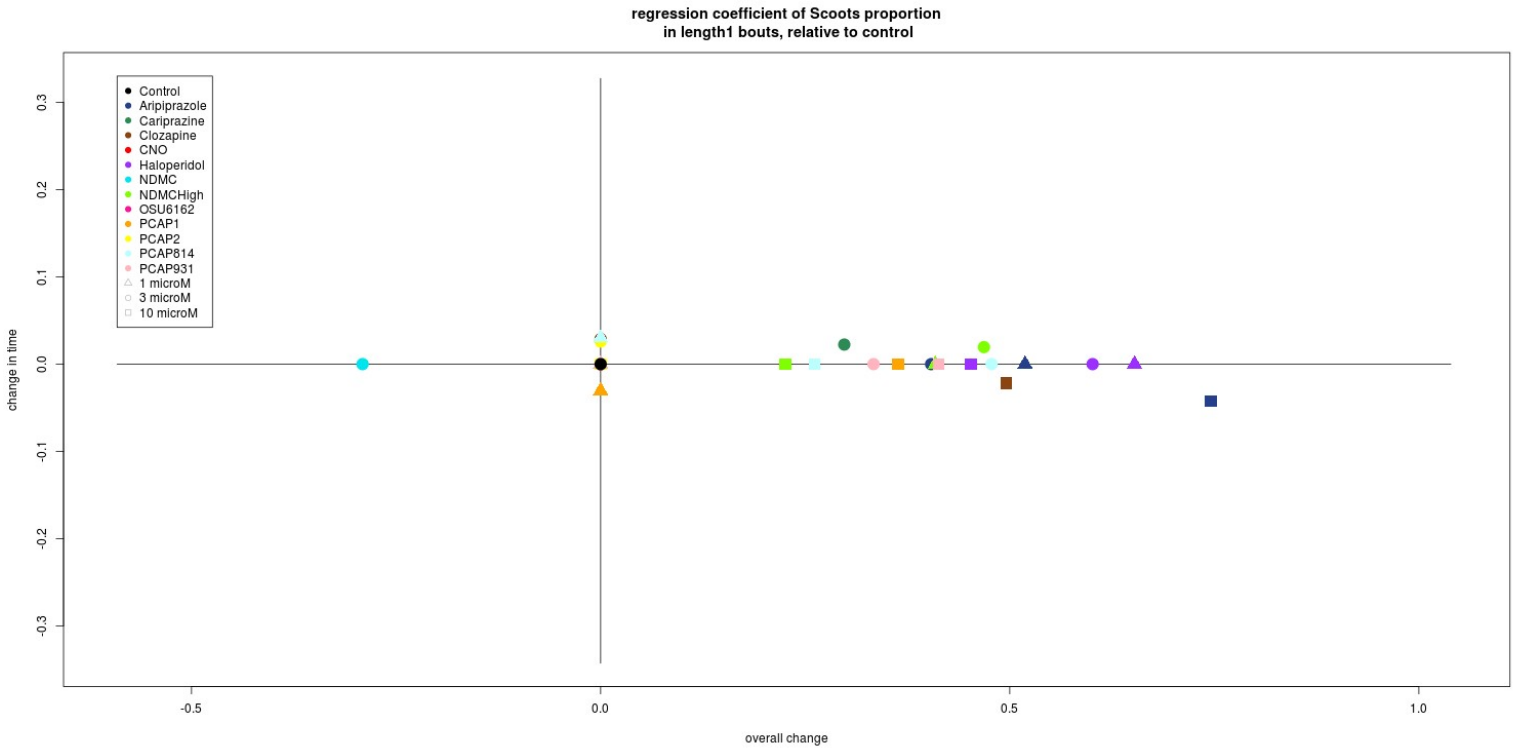
DarkPTZ, permille of length1 bouts:



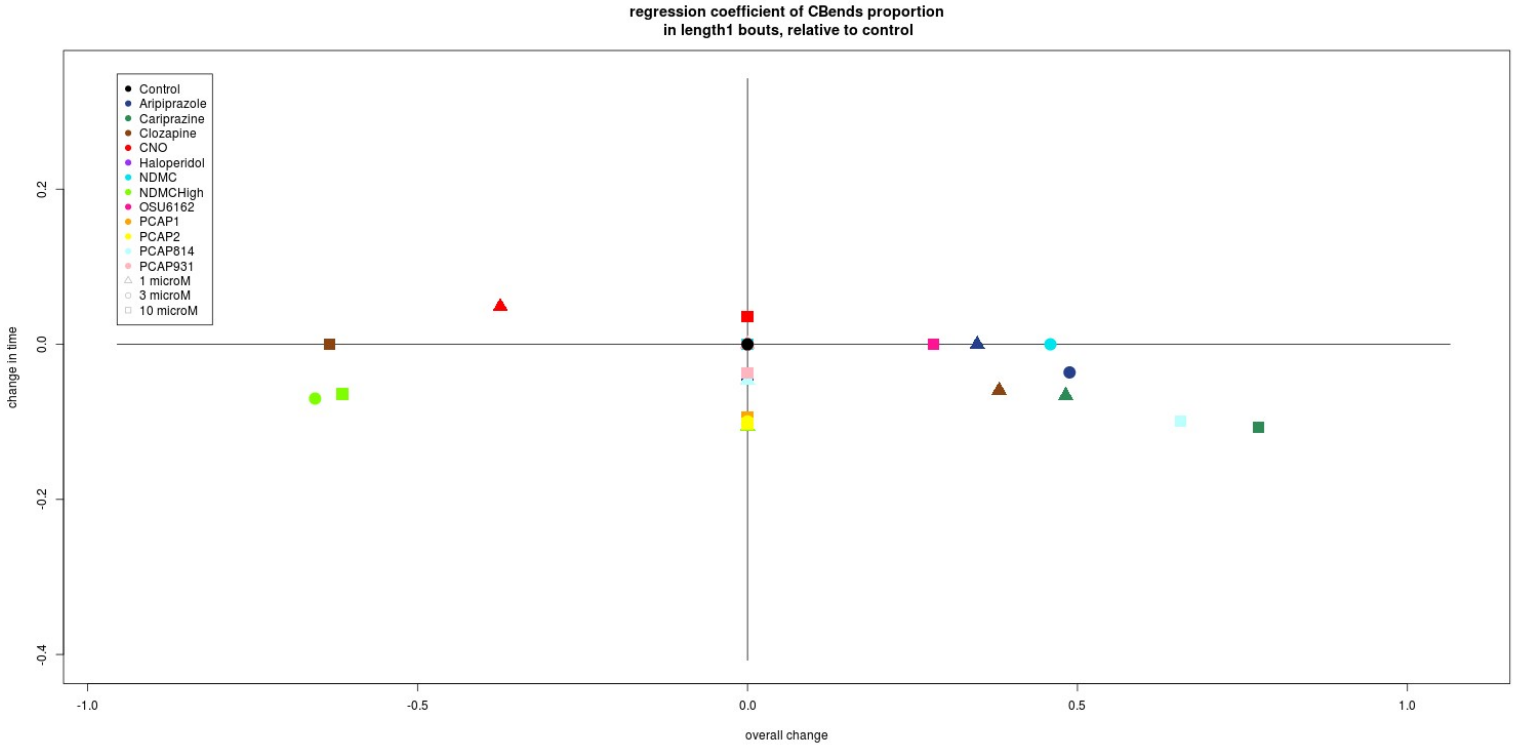
Dark, proportion of scoots in length 1 bouts:



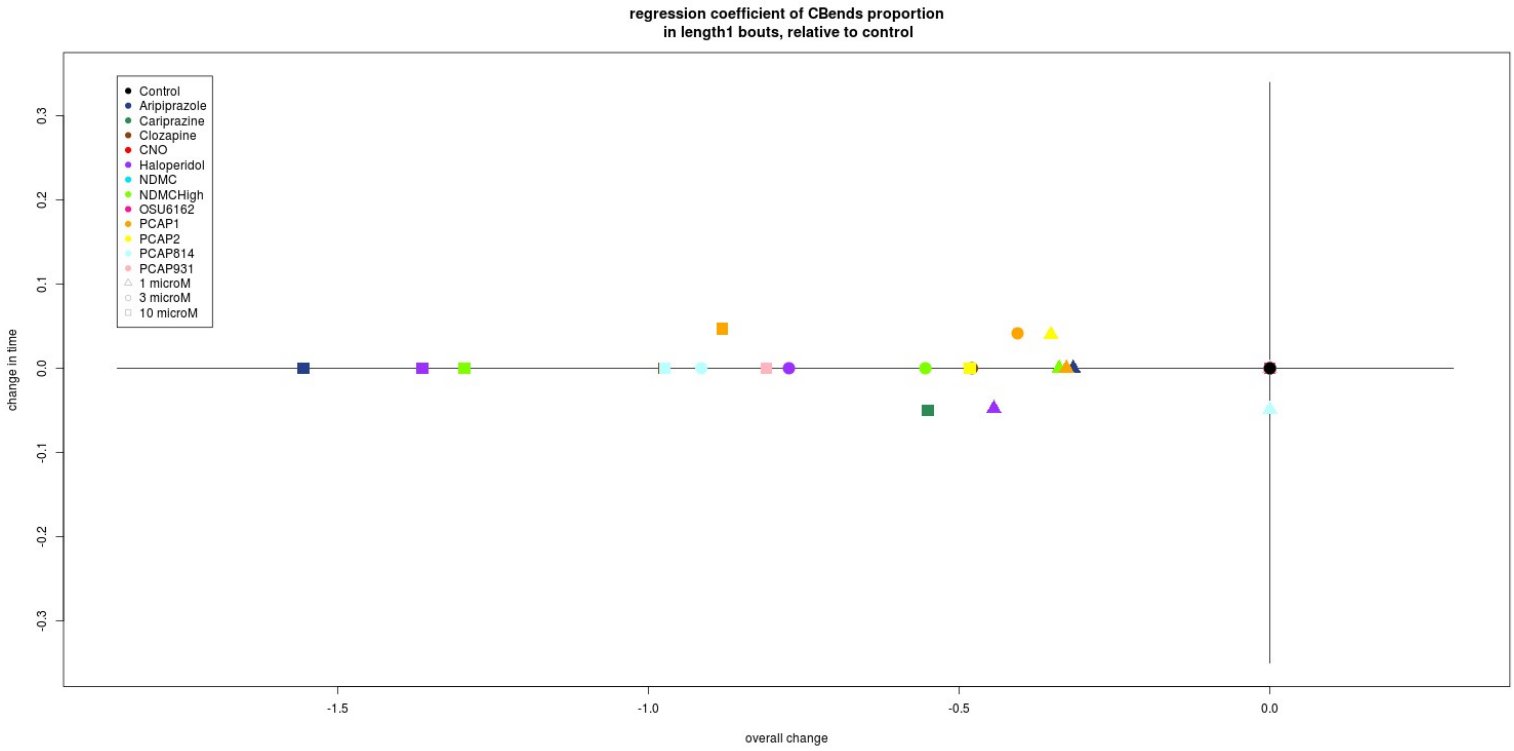
DarkPTZ, proportion of scoots in length 1 bouts:



Dark, proportion of CBends in length 1 bouts:



DarkPTZ, proportion of CBends in length 1 bouts:



If the models and results interpretations are ok I will try to discuss with Anna if it would be ok to extend to model for all bout lengths together, either having variables from stratified bouts by length or adjust for length and then do variable selection including all variables, from counts, bout lengths, turn proportions, transitions etc…
This could then be used as the final analysis, comparing different drugs.