

# Pilot Study

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(gt)
```

## Aims

The aims of this search validation is firstly to determine if the number of regular expression (regex) matches associated with each paper can be used to identify publications that are of higher priority for screening for this project due to the limited time frame. Secondly, it aims to assess the two different types of regex matches, “full text” where the software screens the text as a whole and “Title Abstract” (TIAB) which only looks at the title and abstract in order to determine which type of screening is more appropriate.

## Methods

### Study Identification

Publications contained in the AD-SOLES (<https://camarades.shinyapps.io/AD-SOLES/>) database (n = 32980) were full text and TIAB screened using the following regular expression (regex) term:

In-vitro electrophysiology: S[slice].[pP]reparation|[Hh]ippocampal.slice.|[tT]ransverse.vibratome.sections|[tT]ransverse.slice.|[Ee]lectrophysi.|[fF]ield.(excitatory|potential|postsynaptic)|fEPSP.|[Ff]ield EPSP.

This regex term was developed in order to find studies which contain hippocampal slice electrophysiology outcomes. Full text screening and TIAB screening were both carried out of papers published from 2018-2023 as this is an update of a project that included all papers prior to 2018.

### Inclusion and Exclusion Criteria

A random sample of these papers were reviewed manually for inclusion or exclusion using the System Review Facility (SyRF). Inclusion criteria; Primary experiments in transgenic animals of any sex, age, or species. All languages were included Exclusion criteria: Primary experiments in non-transgenic or combined AD animal models, or studies in humans, in vitro, ex vivo, or in silico. Screening was carried out by 2 reviewers.

## Results

### Random Sampling of Data Set

The regex screening of the AD-SOLES database found n = 2109 publications. A random sample of n = 100 of the 2109 results were uploaded to the SyRF (System Review Facility) project.

```

citations <- read_csv("citations-2023-01-20.csv") # n = 2109

citations_sample <- citations %>%
  select(Title = title,
         Authors = author,
         PublicationName = journal,
         Abstract = abstract,
         Year = year,
         Doi = doi,
         Url = url,
         KeyWords = keywords,
         CustomId = uid) %>%
  mutate(AlternateName = "",
         AuthorAddress = "",
         ReferenceType = "",
         PdfRelativePath = "") %>% # creating correct columns for SyRF system
  sample_n(100) %>% # selecting 100 random papers from data set
  #write_csv("PilotSamplePapers.csv") # saves into dataset

```

## Expanding the Data Set

Due to the exploratory nature of this pilot study, after the selection and manual screening and analysis of the initial 100 random publications which was found to have sensitivity of 96.7% > 1 regex match and specificity of 45.8% > 1 regex match (calculations not included), it was decided that the study would be expanded to 200 papers so our sample size was ~10% of the data set and our calculations could be powered appropriately.

The original 100 papers were filtered out of the data set and an additional 100 papers selected at random were manually filtered for inclusion or exclusion.

```

screening_1 <- read_csv("Screening_Data.csv") # initial 100 papers

filtered <- screening_1 %>%
  mutate(IndicatorColumn = "Yes") %>%
  select(title = Title, IndicatorColumn) %>%
  full_join(citations) %>%
  filter(is.na(IndicatorColumn)) %>% #filtering out initial 100 papers
  sample_n(100) %>% #selecting new 100 random papers
  select(-IndicatorColumn)

```

## Full Text Regex Matches

Manual screening decisions of the for inclusion and exclusion was uploaded and plotted against regex count. Regex count refers to the number of times the software picked up the “regular expressions” searched for in the full text of the paper.

```

# downloading screened data from SyRF
screening_2 <- read_csv("Screening_Data_3.csv") # screening data from 200

regex_matches <- read_csv("regex_matches.csv") %>%
  rename(CustomId = "uid") #regex matches associated with each paper
# Join regex matches to the screening data
screening_regex_fulltext <- screening_2 %>%

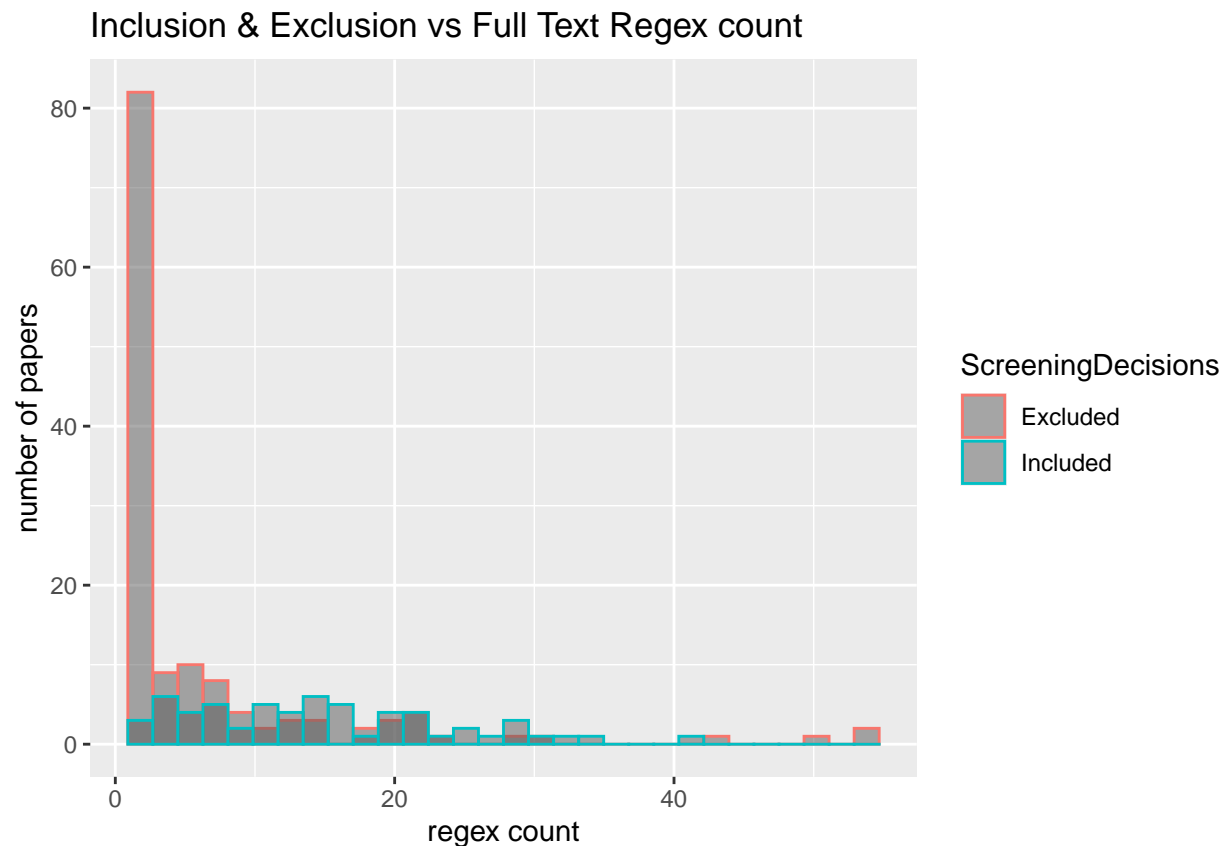
```

```

left_join(regex_matches) %>%
  filter(method == "full_text") %>% #filter out the tiab regex matches
  select(ScreeningDecisions, regex_count)

#plot histogram
histogram <- screening_regex_fulltext %>%
  ggplot(aes(regex_count, colour = ScreeningDecisions)) +
  geom_histogram(alpha = 0.5, position = 'identity')+
  labs(title= "Inclusion & Exclusion vs Full Text Regex count",
       x= "regex count",
       y= "number of papers")
histogram

```



## Sensitivity and Specificity of Full Text

The sensitivity and the specificity of inclusion of papers with a regex count of >1 calculated, as well as the true positives, false positives, true negatives and false negatives;

sensitivity = number of true positives/ number of true positives + number of false negatives

specificity = number of true negatives/ number of true negatives + number of false positives

```

all_positive_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Included")
all_positive_fulltext <- nrow(all_positive_fulltext)

```

```

false_negatives_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Included" & regex_count < 2)
false_negatives_fulltext <- nrow(false_negatives_fulltext)

true_positives_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Included" & regex_count > 1)
true_positives_fulltext <- nrow(true_positives_fulltext)

all_negatives_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Excluded")
all_negatives_fulltext <- nrow(all_negatives_fulltext)

true_negatives_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Excluded" & regex_count < 2)
true_negatives_fulltext <- nrow(true_negatives_fulltext)

false_positives_fulltext <- screening_regex_fulltext %>%
  filter(ScreeningDecisions == "Excluded" & regex_count > 1)
false_positives_fulltext <- nrow(false_positives_fulltext)

sensitivity_fulltext <- true_positives_fulltext/(true_positives_fulltext +
  false_negatives_fulltext) * 100

specificity_fulltext <- true_negatives_fulltext/(true_negatives_fulltext +
  false_positives_fulltext) * 100

```

The definitions for true positives, false positives, true negatives and false negatives are as follows;

- true positive = a study that was included at manual screening and correctly included with a regex count cut-off at >1
- true negative = a study that should be excluded at manual screening is excluded correctly with the regex count cut-off of >1
- false positives = a study that should be excluded at screening is wrongly included with the regex count cut-off of >1
- false negatives = a study that should be included at manual screening is wrongly excluded with the regex count cut-off of >1

## Title & Abstract Regex Matches

```

screening_regex_tiab <- screening_2 %>%
  left_join(regex_matches) %>%
  filter(method == "tiabkw") %>% #filter out the tiab regex matches
  select(ScreeningDecisions, regex_count)

#plot histogram
histogram <- screening_regex_tiab %>%
  ggplot(aes(regex_count, colour = ScreeningDecisions)) +
  geom_histogram(alpha = 0.5, position = 'identity') +
  labs(title= "Inclusion & Exclusion vs Title Abstract Regex Count",

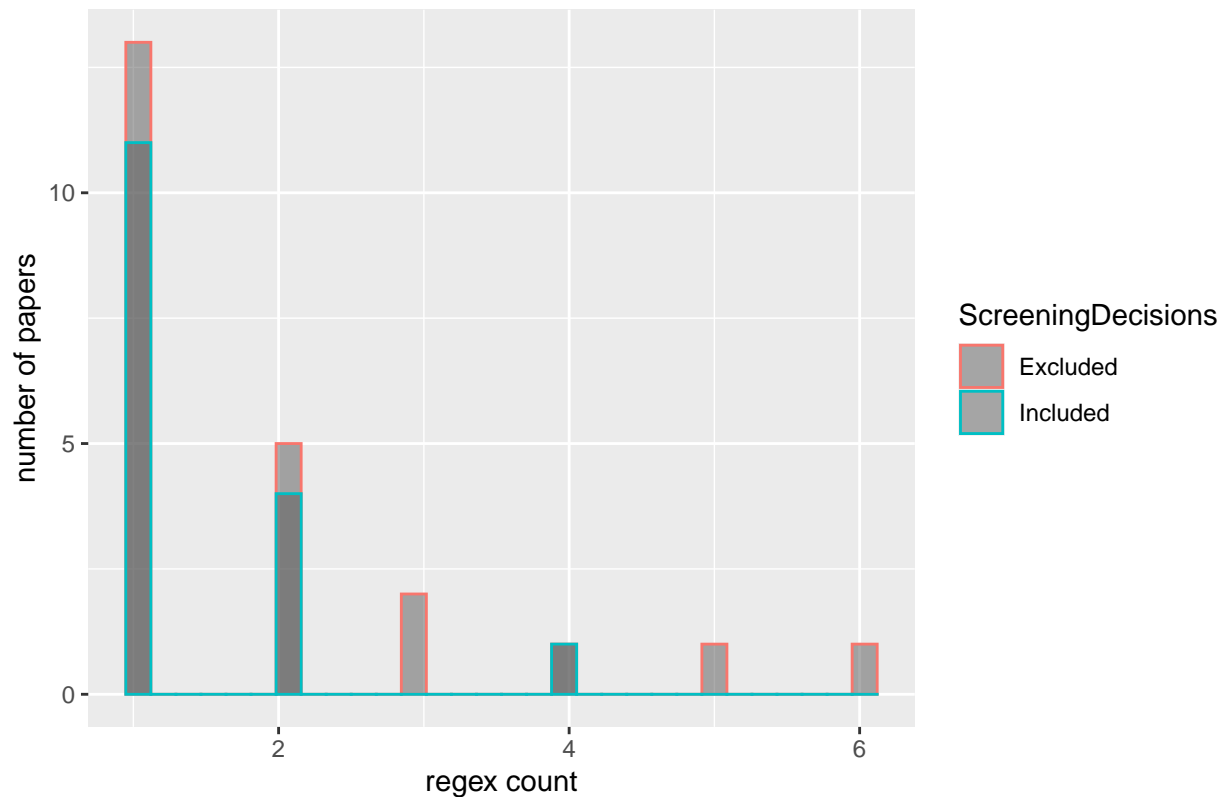
```

```

x= "regex count",
y= "number of papers")
histogram

```

## Inclusion & Exclusion vs Title Abstract Regex Count



## Sensitivity and Specificity

```

all_positive_tiab <- screening_regex_tiab %>%
  filter(ScreeningDecisions == "Included")
all_positive_tiab <- nrow(all_positive_tiab)

false_negatives_tiab <- screening_regex_tiab %>%
  filter(ScreeningDecisions == "Included" & regex_count < 2)
false_negatives_tiab <- nrow(false_negatives_tiab)

true_positives_tiab <- screening_regex_tiab %>%
  filter(ScreeningDecisions == "Included" & regex_count > 1)
true_positives_tiab <- nrow(true_positives_tiab)

all_negatives_tiab <- screening_regex_tiab %>%
  filter(ScreeningDecisions == "Excluded")
all_negatives_tiab <- nrow(all_negatives_tiab)

true_negatives_tiab <- screening_regex_tiab %>%

```

```

    filter(ScreeningDecisions == "Excluded" & regex_count < 2)
true_negatives_tiab <- nrow(true_negatives_tiab)

false_positives_tiab <- screening_regex_tiab %>%
  filter(ScreeningDecisions == "Excluded" & regex_count > 1)
false_positives_tiab <- nrow(false_positives_tiab)

sensitivity_tiab <- true_positives_tiab/(true_positives_tiab +
  false_negatives_tiab) * 100

specificity_tiab <- true_negatives_tiab/(true_negatives_tiab +
  false_positives_tiab) * 100

```

## Summary of Results

```

results <- data.frame(Screening = c("Full Text", "Title Abstract"),
  TruePositives = c(true_positives_fulltext, true_positives_tiab),
  TrueNegatives = c(true_negatives_fulltext, true_negatives_tiab),
  FalsePositives = c(false_positives_fulltext, false_positives_tiab),
  FalseNegatives = c(false_negatives_fulltext, false_negatives_tiab),
  Sensitivity = c(sensitivity_fulltext, sensitivity_tiab),
  Specificity = c(specificity_fulltext, specificity_tiab))

table <- results %>%
  gt() %>%
  tab_header(title = "Summary of Pilot Study Results")

table

```

Summary of Pilot Study Results

Screening	TruePositives	TrueNegatives	FalsePositives	FalseNegatives	Sensitivity	Specificity
Full Text	57	60	77	3	95.00	43.79562
Title Abstract	5	13	10	11	31.25	56.52174

## Conclusions

The results show that it is possible to use full text regex terms with a high degree of sensitivity, meaning that 95% of papers that should be included in the study are found in papers that have a full text regex count of >1. The comparison between the full text regex count and the title abstract count shows that it is better to use full text screening for identification of studies as it is much more accurate. This is because electrophysiology is an outcome that is not usually mentioned in the title or abstracts of papers. This shows that this method of paper identification for the present systematic search is much more useful for finding relevant studies with specific outputs.