EMS: Biomedical Sciences
University of Edinburgh
Honours Project

Neuroscience
2022-23

B152389

A Systematic Review and Meta-Analysis of Synaptic Plasticity in APP Models of Alzheimer's Disease

Analysis Project - Systematic Review
Word Count: 4955

BIOMEDICAL TEACHING ORGANISATION

Retention of course work form

With your consent and with your name removed, your dissertation may be retained by the Deanery of Biomedical Sciences for 10 years and made available to other students. Note however that we cannot anonymise video recordings. This collection of work is held to assist future students and you may have used this resource in planning your own projects. If you do not consent to your work being retained, please specify this below and we will destroy the work in accordance with Taught Assessment Regulation 49.

https://www.ed.ac.uk/files/atoms/files/taughtassessmentregulations.pdf

If in future you wish to withdraw consent, please email bmto@ed.ac.uk

Please select one of the following options

I consent to my work being retained and made available to future students.	X
I do NOT consent to my work being retained and made available to future students.	
My work contains video recording and I understand that it might be retained and made available to future students. I consent to this with full understanding that it cannot be anonymized.	

Contents

Acknowledgements

I would like to say a huge thank you to my supervisor Dr Kaitlyn Hair who was such an incredible mentor and teacher through this whole project. I would also like to say thank you to everybody on the CAMRADES research team who supported us with our research.

Abbreviations

Abbreviation	Meaning
AD	Alzheimer's Disease
Aβ	Beta Amyloid
APP	Amyloid Precurser Protein
PSN1	Presenilin-1
NFTs	Neurofibrillary Tangles
LTP	Long term potentiation
LTD	Long term depression
PPF	Paired pulse frequency
PPR	Paired pulse ratio
I/O	Input output
MWM	Morris water maze
regex	Regualr Expressions
SyRF	Systematic Reveiw Facility
SMD	Standard mean difference
SD	Standard deviation
SEM	Standard error measure
REMIL	Restricted maximum likelihood

Table 1: *List of abbreviations used*

Abstract

Alzheimer's disease (AD) in a progressive neurodegenerative disorder for which there is currently no effective disease altering treatment available. Most research in this area is aimed at reducing levels of amyloid beta build-up in the brain however despite extensive research, only a single treatment has been licensed. This has lead researchers to look at other targets, such as synaptic dysfunction which occurs earlier in disease progression. AD pre-clinical research utilises transgenic animal models, but the extent to which these models show synaptic dysfunction has shown varied results. This study used systematic review and meta-analysis techniques to analyse the impact of modelling AD with mutations in amyloid precursor protein on synaptic plasticity outcomes and identify sources of heterogeneity across publications. Results were pooled from 24 papers screened from 2109 studies. Meta analysis showed synaptic plasticity outcomes were found to be reduced in animal models of AD but it was not possible to identify sources of heterogeneity in this analysis. This study established the extent of the modelling effect of APP models of AD which giving a better understanding of baseline synaptic function, which would be a highly useful tool for the drug development pipeline.
word count:193

Lay Abstract

Alzheimer's disease (AD) causes the brain to start to loose mass, impacting function and causing issues with memory which get worse over time. The only treatments available to people currently treat the symptoms of the disease and not the underlying cause, with only one drug like this licenced for use. In order to develop more treatments, scientists think they are able to target the problem with the connection between brain cells which occurs very early on in the disease. Studies investigating this have produced conflicting results at the animal research stage. To understand the true extent of the problems in brain cell connections in animals used to research AD and to identify reasons that the results were so varied across studies, results from all studies were collected and analysed. This showed that the animals used to model AD did show a significant problem in the connection between brain cells early on in disease compared to normal animals but was not able to identify the reason for the varied results. This established a baseline for the brain cell connection issues in animals used to research AD which is useful for drug development purposes.
word count:192

1 Introduction

1.1 Alzheimer's Disease

Alzheimer's Disease (AD) is a multifactorial progressive neurodegenerative disorder that causes loss of cognitive function and severe memory impairment, impacting language, executive function, and visual-spatial abilities. It is responsible for 60-80% of Dementia cases, a disease that currently effects over 55 million people (Organisation 2021) with rates projected to triple worldwide by 2050 (Marchetti & Marie 2011). As it stands, there is not a complete understanding of the underlying neuropathology of AD, however pathological hallmarks of the disease are extracellular β-amyloid (Aβ) buildup, intracellular tau neurofibrillary tangles (NFTs), glial cell dysfunction, synaptic loss, and neuroinflammation (Scheltens 2021). The amyloid hypothesis is a widely studied theory that proposes that the accumulation of Aβ is the primary driver of AD pathology (Hardy & Higgins 1992)(Figure1), however a full understanding of the mechanisms has not yet been achieved.

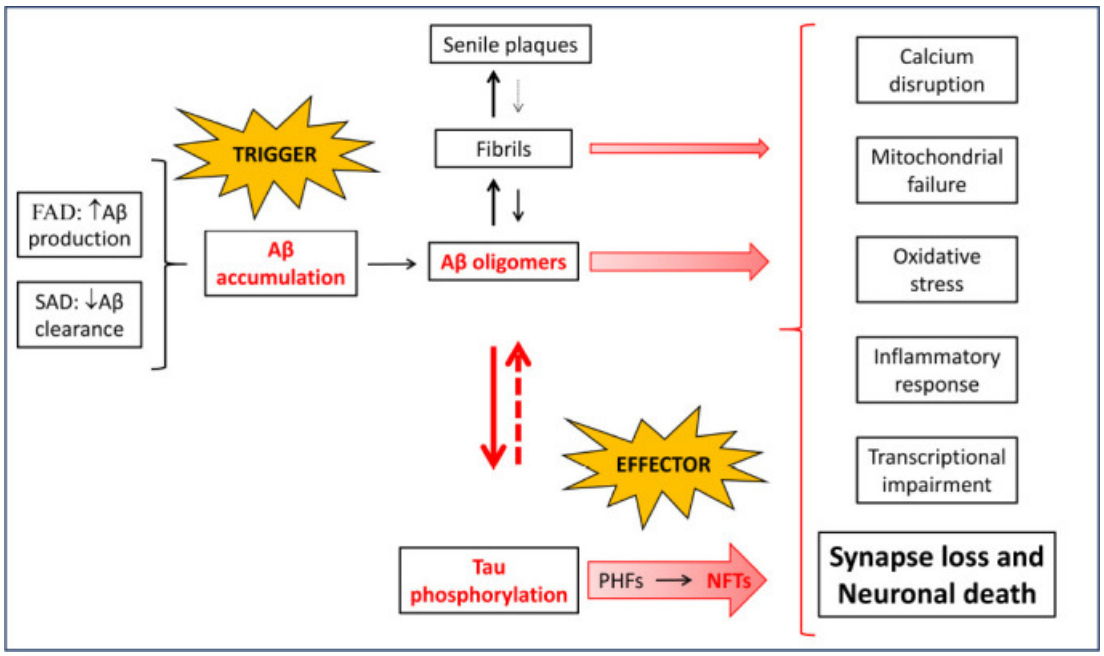


Figure 1: *Amyloid Cascade Hypothesis showing the proposed relationship between Aβ and NFTs leading to neurodegeneration Villegas et al. (2022)*

1.2 Treatment

After decades of research targeting the reduction of Aβ levels, a disease altering treatment was licensed for use for the first time in 2021 (Golde 2022). Some suggest that the difficulty producing an effective disease altering treatment targeting Aβ is a result of the disease progression being too advanced by the time Aβ buildup has occurred . This has lead researchers to explore other therapeutic targets (Sri et al. 2019). There is evidence of synaptic dysfunction occurring in the hippocampus before advanced stages of Aβ plaque development and general neuronal loss in humans (Marchetti & Marie 2011). This suggests AD could be a result of synaptic failure, supported by the growing body of evidence showing

that the extent of synapse loss seen in diseased brains is the strongest correlate to the cognitive impairment observed in AD patients (D'Amelio et al. 2010). This research suggests there may be therapeutic potential in targeting synaptic dysfunction in in the early stages of the disease (Koffie et al. 2011).

1.3 Synaptic Plasticity

Synaptic plasticity refers to the activity-dependent change of the strength of synaptic transmission at synapses and is a widely accepted model of learning memory formation (Egan et al. 2016). Synapse dysfunction is often studied using a measure of synaptic plasticity in the form of long term potentiation (LTP) and depression (LTD). These are the long lasting activity based modifications of synaptic strength, with LTP conceptualising a long lasting increase in synaptic transmission after stimulation to the excitatory pathways and LTD the weakening of synapses by infrequent stimulation (Citri & Malenka 2007). These measures are often used alongside the Morris water maze (MWM) paradigm that assesses learning and memory directly from animal behaviour (Othman et al. 2022). In pre-clinical research, synaptic plasticity is often assessed by measuring field potentials in *in vitro* hippocampal slices derived from rodent models. LTP can be induced by applying either high frequency stimulation or theta burst stimulation and LTD can be induced by low frequemcy stimulation. To measure changes in LTP/ LTD, field excitatory post-synaptic potentials (fEPSPs) are recorded to determine a baseline before the induction protocol is applied. To determine basal synaptic transmission and to calculate appropriate baseline stimulation voltages for LTP/LTD experiments, simulations of increasing voltages are applied, and the resulting output is measured at each step and plotted as the input / output (I/O) relationship. The techniques are a useful tool for building a better understanding of synaptic dysfunction for the drug development pipeline.

1.4 Transgenic Animal Models of AD

Transgenic animal models are widely used for pre-clinical research in AD including investigations into synaptic dysfunction. Transgenic modelling involves altering the genome of animals to include a transgene or foreign sequence in order to study the gene function and model disease (Elder et al. 2010). There have been more than 40 Alzheimer's disease-associated genetic risk loci identified in humans giving a variety of avenues for modelling in animals (Scheltens 2021). Transgenic modeling in not able to produce a model displaying full typical AD pathology as 95% of AD cases are sporadic and have no known genetic link, however it allows for the investigation of specific known AD pathology's (Elder et al. 2010). Mutations in the amyloid precursor protein (APP) have been used to produce a number of mouse models including Tg2576, which over expresses a mutant APP with the Swedish mutation and develops extensive amyloid plaque physiology at a young age (Hsiao et al. 1996, Mucke et al. 2000).

1.5 Research to Date

If synaptic dysfunction is to be targeted by the drug pipeline, a better understanding of synaptic plasticity deficits in models of AD is needed as experimental results so far have varied (Chapman et al. 1999, Fitzjohn et al. 2001, Marchetti & Marie 2011). A systematic review carried out by Hair (2021) previously investigated *in vitro* field electrophysiolgy in hippocampal slices across multiple transgenic APP models of

AD. The review found heterogeneity between studies, with some studies showing an increase in synaptic function between the model and wild-type, and others showing a deficit. The analysis also detected publication bias across the literature. This project is an update of (Hair 2021) systematic review aimed at analysing and synthesising the data produced in the literature since 2018.

1.6 Research Questions

The research questions for this study are as follows:

- What is the impact of transgenic modelling of AD with mutations in the APP gene on synaptic plasticity outcomes?
- What is the prevalence of risk of bias reporting in the literature and what is the impact of this on effect sizes?
- What experimental design variables are reported in *in vitro* hippocampal slice electrophysiology studies and what is the influence of these factors?

These will be explored using systematic review and meta-analysis techniques.

2 Methods

The protocol detailing the aims and methodology was preregistered on the Open Science Framework (Appendix 1). The protocol was updated shortly after the screening process started to include an additional variable. What plane the hippocampal slice was taken from (e.g. sagittal, coronal or horizontal) was added after identifying this differed between studies and could be introducing variation that had not been considered originally (Xiong et al. 2017).

2.1 Study Identification

Publications measuring synaptic plasticity in hippocampal slices *In vitro* were identified by searching the AD-SOLES database for studies published from 2018-2023. AD-SOLES is a website (https://camarades.shinyapps.io/ad-soles/) containing a living database that systematically collects and synthesises all experimental evidence in animal models of Alzheimer's disease using an integrated series of automated tools including machine-learning approaches to continually include new evidence published (Hair et al. 2022). It uses regular expressions (regex), which are a sequence of characters used to define a search pattern, applied to the title and abstract and the full text to tag papers with outcome measures. The regex term used to to identify papers with electrophysiology outcomes is included in Appendix 2. The AD-SOLES database has built in de-duplication, however a further de-duplicaiton step was carried out using excel to make sure there was no overlap with the previous systematic review which included some papers from 2018 (Hair 2021).

2.2 Pilot Study

A pilot study (Appendix 3) was carried out in order to identify studies that were of higher priority for screening due to the limited time frame of this project. After study identification, a random selection of approximately 10% of publications identified were manually screened for inclusion and exclusion. The inclusion / exclusion criteria for the pilot study are as follows:

- 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.

The sensitivity and specificity of of excluding papers with a threshold of >1 regex was calculated using Equations (1) and (2).

$$\text{Sensitivity} = \frac{\text{Number of True Positives}}{\text{Number of True Positives + Number of False Negatives}} \tag{1}$$

$$\text{Specificity} = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives + Number of False Positives}} \tag{2}$$

Where;

- 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

It was determined that of all relevant papers within the data set, 95% are found in publications that have a full text regex match of >1 (Table 5), meaning these papers were prioritised for screening in this project and publications that did not meet this threshold were excluded.

Screening	True Positive	True Negative	False Positives	False Negatives	Sensitivity	Specificity
Full Text	57	60	77	3	95.00	43.80
Title Abstract	5	13	10	11	31.25	56.52

Table 2: *Summary of Pilot Study Results*

2.3 Study Selection Criteria

Studies measuring synaptic plasticity or transmission using field potentials in hippocampal slices from transgenic animal models of AD with mutations in the Amyloid Precursor Protein (APP) only were included, with no restrictions on developmental stage. The narrowed scope to include only one model type was due to the limited time frame of this project. Models with mutations in the APP gene were chosen as these have historically been one of the most commonly used for research due to the Aβ pathology typically seen (Hair 2021). The control population was defined as a cohort of wild-type animals from the same litter as transgenic animal models or an age-matched wild-type with the same background strain.

Variable	Inclusion Criteria	Exclusion Criteria
Study Design	All primary experiments in transgenic AD animal models	Studies without a proper control and non-primary research including reviews, commentaries, letters, and editorials.
Animal Model	Transgenic AD animal models with APP mutations of any age, sex, or species.	Primary experiments in non-transgenic or combined AD animal models, human studies, in vitro, ex vivo, or in silico studies.
Outcome Measures	Studies which measure field synaptic transmission (I/O relationships) or plasticity (LTP/LTD/PPF) using in vitro electrophysiology in hippocampal brain slices. Behavioural data from Morris water maze experiments will also be included	Studies which do not contain electrophysiology outcomes, measure alternative brain areas, or use only whole-cell techniques to measure intracellular potentials.
Language	All languages	None
Publication Date	2018-2023	Studies published before 2018

Table 3: *Inclusion and exclusion criteria*

2.4 Data Extraction

The details of the variables extracted from relevant publications can be found in Table 4. Whether number of slices or animals were reported as the unit of analysis for statistical tests was also extracted, as using slices as the experimental unit in these studies risks introducing dependencies between data points and is a form of pseudoreplication. Studies were extracted regardless of sample size reporting, however only studies that stated that the sample size was equal to the number of animals were included in the meta analysis.

Information Extracted	Additional Information
Study Meta-Data	
Title	
DOI	
First Author	
Corresponding journal name	
Country of origin of corresponding author	

Information Extracted	Additional Information
Measures to Assess Risk of bias	
Reporting of random allocation of animals to experimental groups	
Reporting of blinding during outcome assessment	
Reporting of sample size calculation	
Reporting of study protocol dated before study began	
Reporting of potential conflicts of interest statement	
Reporting of rationale for data point inclusion/ exclusion	
Animal Husbandry	
Light cycle of animal facilities	Light cycle hours (e.g 12)
Number of animals per cage	
Environmental enrichment reported	E.g toys in cages
Model Induction	
Animal species	
Animal background strain	
Source of animals	
Transgenic animal model	
The type of model control	E.g. WT littermates or age matched same background strain
Age of animals at time of outcome assessment	Mean calculated when given as a range
Sex of animal	
Outcome Measures	
Outcome measure average type	
Outcome measure error type (SD or SEM)	
Brain pathway recorded	Schaffer collaterals CA1 - CA3 or Dentate Gyrus
Outcome measure type	LTP, LTD, PPF, I/O, MWM
Outcome measure units	
Outcome measure detail	E.g % Normalised fEPSP slope
Time taken of baseline recording	
Blockers used while investigation synaptic plasticity/ transmission	
Time of day sacrifice	
Anaesthesia used prior to decapitation	
Plane of brain slice	
Percentage of maximal response from I/O used for baseline recordings	
Calcium concentration of recording solution	

Information Extracted	Additional Information
Magnesium concentration of recording solution	
Calcium concentration of slicing solution	
Magnesium concentration of slicing solution Magnesium concentration of slicing solution	
Presence of kynurenic acide in slicing solution	
Length of time brain slices left to recover prior to stimulation	
Temperature slices left to recover at	
Type of recording chamber	Submersion, interface chamber or multi-electrode array
Electrophysiology Protocols	
Type of stimulation used to induce LTP/LTD	
Number of stimulation trains	
Inter-stimulation interval	
Number of bursts in 1 stimulation train	
Frequency (Hz) of bursts in 1 stimulation train	
Inter-burst interval in 1 stimulation train	
Inter-burst stimulation units	
Number of pulses in 1 burst	
Frequency (Hz) of pulses in 1 burst	
Inter-pulse interval in 1 burst	
Inter-pulse interval units	
Percentage (%) of maximal response used for induction	
Total number of stimulations	
Morris Water Maze Paradigm Protocols	
Type of outcome	acquisition vs. probe
Number of training days	
Number if trials per day	
Whether a visible platform test was used	
Pool temperature	
Acquisition/ probe phase outcome measure unit	E.g. escape latency (s)/ training days

Table 4: *Information extracted from publications*

Publications identified as relevant for this project were uploaded to SyRF. These papers were then screened against the inclusion / exclusion criteria by a single reviewer. Data extraction was completed using SyRF and when outcome data was presented in graphical format, Web Plot Digitiser (https://apps.automeris.io/wpd/) was used to measure means and error bars. Errors detected during data cleaning were corrected in SyRF prior to analysis.

2.5 Random Effects Meta Analysis

Each electrophysiology outcome (LTP, LTD, PPF, I/O) measured in different hippocampal pathways was analysed separately when there was sufficient data (n $>$10). When extracting data about electorphysiology protocols, where LTP or LDP were measured, data point measurements were taken every 30 minutes. To analyse data pooling I/O outcomes across publications, the maximal I/O value which, is typically where the largest difference in basal synaptic strength can be observed between cohorts, was extracted for each experiment. For PPF the inter-stimulus intervals were extracted with the aim to choose the most commonly reported one for analysis however there was not enough data when the final meta analysis was carried out.

Where data was sufficient, the standardised mean difference (SMD) for each modeling versus WT comparison was calculated using the meta package in RStudio (see Appendix 6) for each outcome measure, which uses Hedge's G effect sizes (Vesterinen et al. 2014) for each modeling comparison. Effect sizes were weighted using the inverse variance method in order to reflect each comparison contribution to the total effect estimate. Where there were instances that a single WT control group served multiple transgenic model groups, the group was divided by the number of modelling groups it served. To combine SMD effect sizes for each outcome, a random-effects model with a restricted maximum likelihood estimate of between-study variance was used. This approach assumes that the true effect size variation between studies is due to differences in study design, populations, and modeling.

2.6 Multi-Variable Meta-regression

A multi-variable regression was conducted using RStudio using the metafor package (see Appendix 6) in order to identify sources of heterogeneity across *in vitro* electrophysiology experiments. A model building validation exercise to fit a multivariable meta-regression model as described by Harrer et al. (2021) was conducted. The proportion of heterogeneity accounted for by each variable was determined by first conducting a univariable regression. Categorical variables with less than 10 in one category and only two meaningful categories such as "Reported" and "Not reported", were excluded as there would not be enough statistical power to detect an effect. If a variable was reported in less than 25 experiments then it was was also excluded.

2.7 Meta-Regression: Synaptic Plasticity and Cognition

The MWM data was split into acquisition and probe phase outcomes and the SMD for each comparison calculated using RStudio using the metafor package. If the same cohort of animals were measured using different outcomes for a phase (e.g. both escape latency and distance to platform in the acquisition stage) a nested effect size for that cohort was calculated. This analysis was carried out only if an electrophysiology dataset had n $\geq$ 10 studies where MWM outcomes were included and reported to have been conducted in the same cohort of animals. A univariable meta-regression using the effect sizes (SMD) as the predictor variable was carried out to understand whether cognitive outcomes explained a significant proportion of the heterogeneity seen across publications.

2.8 Publication Bias

It is not possible for any one test to be used to determine publication bias so a range of tests were carried out. The first two methods used were funnel plots and Egger's Test regression (Egger et al. 1997). Code was developed in RStudio (see Appendix 6) to assess funnel plot asymmetry with the Pustejovsky & Rodgers (2019) modification for the standard error shown in Equation 3.

$$\sqrt{W_i} = \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \tag{3}$$

Where $\sqrt{W_i}$ is equal to the modified version of the standard error and where n_1 = sample size of model group and n_2 = sample size of the control group. This avoids the inflation of false positive results when using Egger's Test to assess funnel plots due to SMDs and SEs not being independent of one another. The final method used was trim and fill, which gives a bias-corrected estimate of the true effect size in order to get a sense of the magnitude of the publication bias for each study (Harrer et al. 2021).

2.9 Visualisation of Results

RStudio was used to plot and produce all diagrams presented in Results as shown in Appendix 5.

3 Results

3.1 Study Identification

The publication search on the AD-SOLES database for electrophysiology outcomes, carried out January 24th 2023, identified 2109 potentially relevant studies published from 2018 - 2023. After the criterion of $>$1 full text regex match was applied, 802 were excluded from screening as part of this project and 14 duplicates were removed leaving 1293 publications in the dataset. After screening, 24 studies describing experiments in APP transgenic models were identified where synaptic plasticity/ transmission was investigated through *In vitro* electrophysiology, comparing APP model groups with an appropriate WT control in hippocampal slices. A prisma flow diagram detailing the search is shown in Figure 2 and a full list of publications identified is included in Appendix 5.

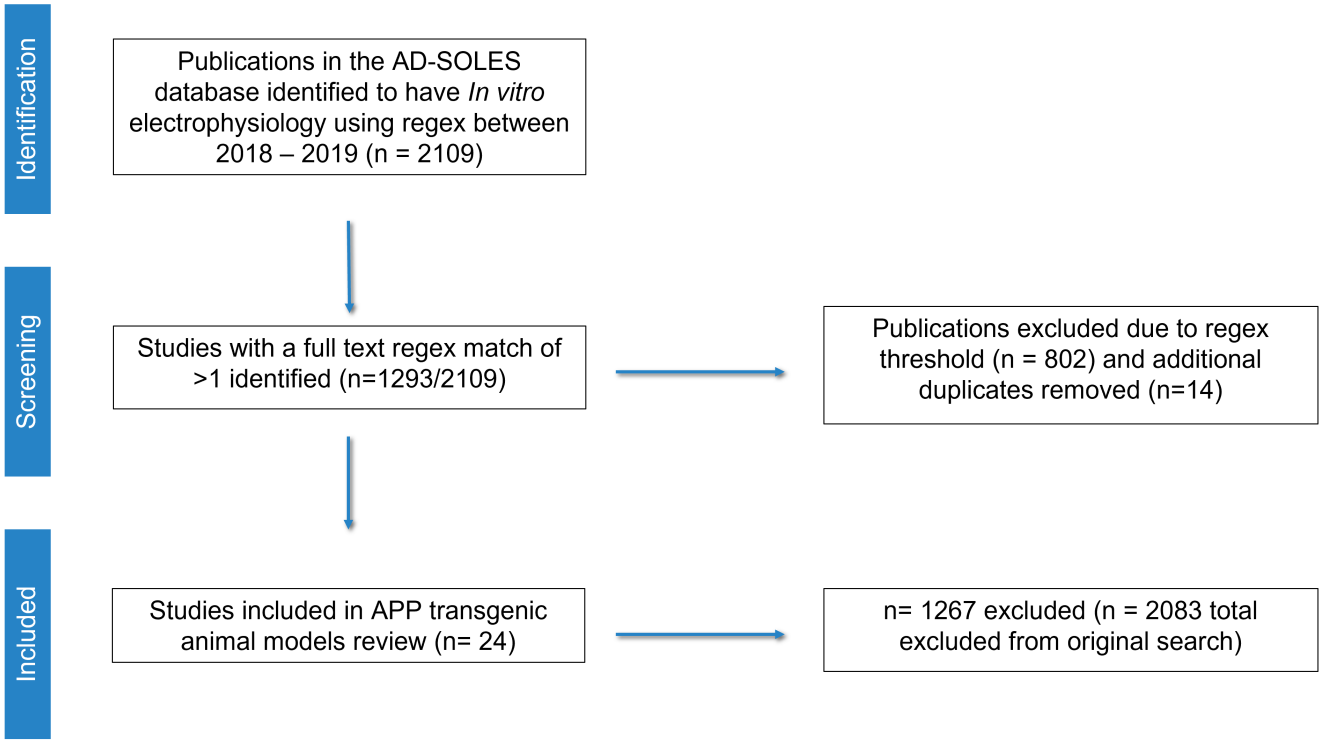


Figure 2: PRISMA Flow Diagram

3.2 Animals

3.2.1 Transgenic APP Models of AD

There were 15 different models used across the 24 studies as shown in Figure 3. Of the fifteen models, twelve models were investigated only once in and three were looked at across multiple studies. The Tg2576 model was investigated in five different studies, while the J20 model was analyzed across four publications, and the APP NL-G-F knock-in model was studied in three different papers.

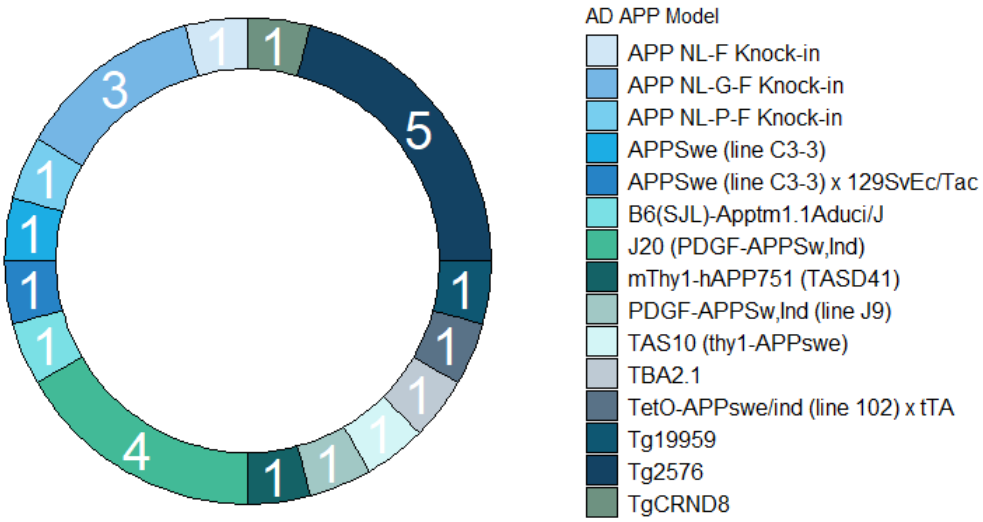


Figure 3: *Different transgenic APP models used across studies (n = 24 publications).*

3.2.2 Age and Sex

Across publications, 76% of experiments used mice less than a year old when investigating synaptic plasticity/transmission (Figure 4). The highest proportion of studies used mice at 24 weeks of age and the oldest mice used were 96 weeks of age.

Around half of studies (48.28%) used mixed groups of male and female mice for experiments. The majority of the remaining studies investigated electrophysiology outcomes in male mice only (46.55%) with a very small number using female mice only (5.17%).

Figure 4: *Distribution and density (shown in red) of age in weeks of animals when electrophysiology outcomes measured (n = 58 experiments).*

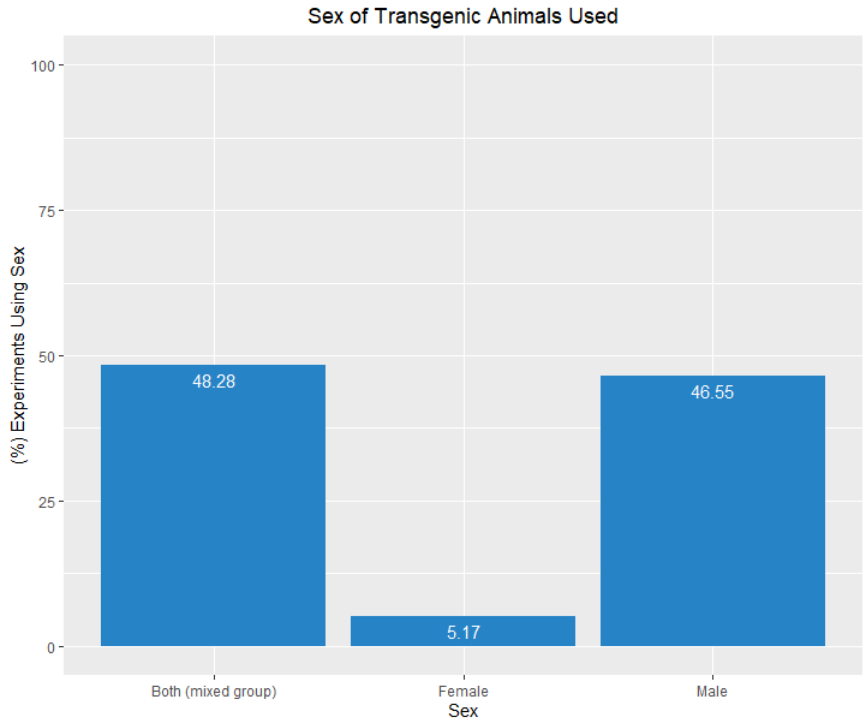


Figure 5: *Proportion of studies using different sex's in their experiments (n = 58 experiments)*

3.3 Reporting Across Publications

3.3.1 Study Quality Protocols

The reporting of quality measures and measures to reduce the risk of bias ranged from well reported to very poorly reported across publications (Figure 6).

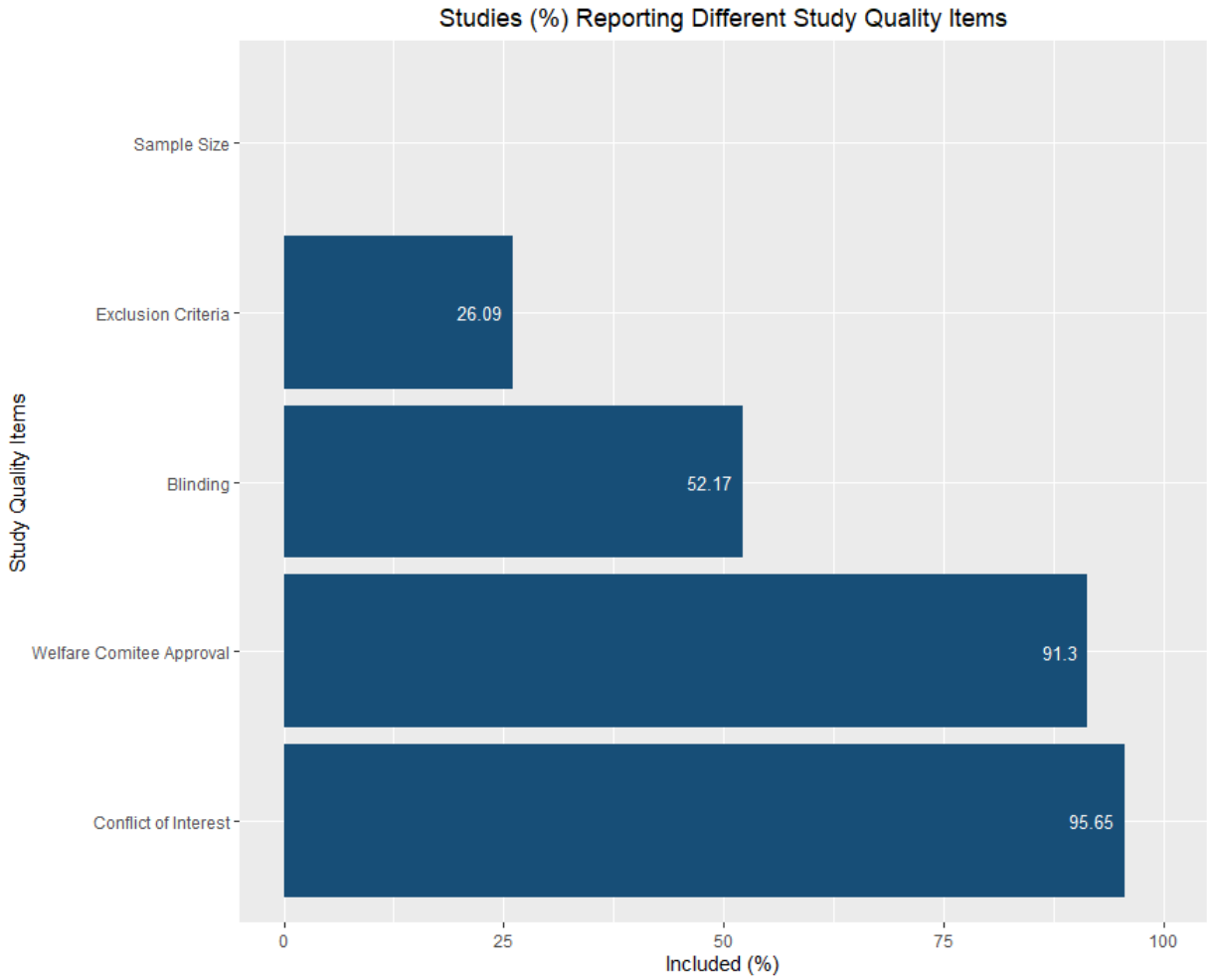


Figure 6: *Barchart showing the reporting of different study quality items.*

It was found that none of the publications reported whether or not a sample size calculation was carried out prior to undertaking the study. Papers determining exclusion criteria for data points (26.09%) and whether experiments were carried out blinded (52.17%) were also very poorly reported. Reporting of animal welfare committee approval and the inclusion of a conflict of interest statement were much more widespread across publications at 91.3% and 95.55% respectively.

3.3.2 Animal Husbandry Protocols

Animal husbandry was poorly to moderately reported across publications, with only 8.7% of studies reporting if the mice experienced any form of environmental enrichment. Less than half (43.47%) reported the number of animals housed per cage and just over two thirds (69.57%) reported the light cycle animals

were housed under (Table 5).

Animal Husbandry Item	% of Studies Reported
Number of Animals per Cage	43.47%
Light Cycle	69.57%
Environmental Enrichment	8.7%

Table 5: *Reporting of animal husbandry protocol measures.*

3.3.3 Slice Preparation Protocols

The reporting of different slice preparation protocol measures again ranged from well reported to very poorly reported, with the use of Kynurnic acid when slicing only reported in 3.7% of studies. Sacrifice time of day (7.41%) and whether or not anaesthesia was used prior to sacrifice (37.04%) were also poorly reported. Slice recovery time (81.48 %), and the concentration of magnesium and calcium in the slicing solution (74.08%) were fairly well reported but the slice protocol measures reported most often were the temperature that hippocampal slices were left to recover at, the plane the brain slice was taken from and the concentration of calcium and magnesium in the recording solution (88.89%).

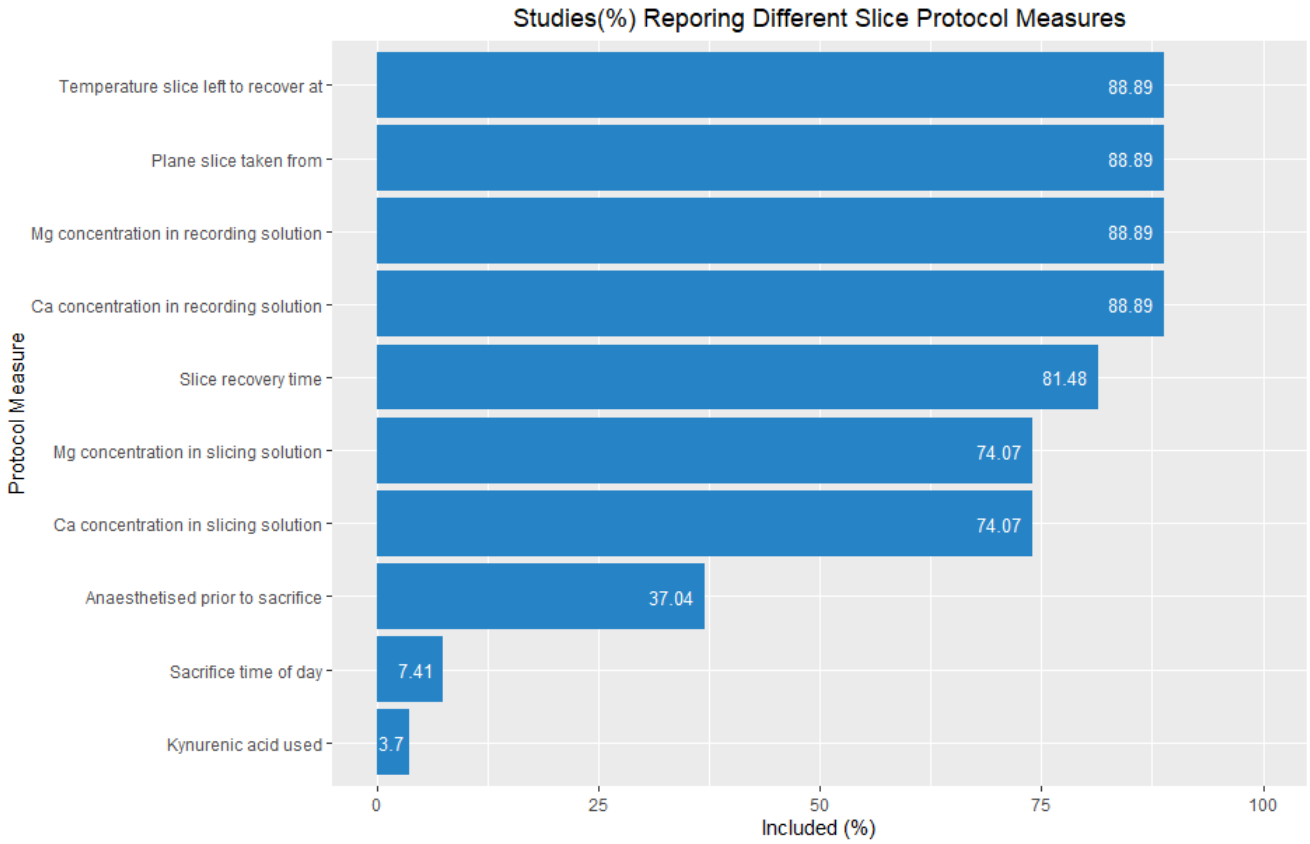


Figure 7: *Reporting of study slice protocol measures.*

3.3.4 Electrophysiology Recording Protocols

The brain pathway recorded was reported in all studies, with all studies assessing the Schaffer Collaterals at the CA1 region. LTP and LTD stimulation type (93%) as well as the number of total stimulation's (89%) was well reported but the percentage of maximal baseline recording (68%) and type of recording chamber (43%) were both included less often.

Electorphysiology Measure	% Experiments Reported*
Brain Pathway	100 %
LTP/LTD Stimulation Type**	93 %
Total Stimulation's**	89 %
% of Maximal of Baseline Recording**	68 %
Type of Recording Chamber	43 %

* 58 experiments across 24 studies assessing synaptic plasticity/ transmission

** Across 28 experiments assessing LTP/LTD

Table 6: *Electrophysiology reporting quality.*

3.4 Random Effects Meta-Analysis

For meta analysis, 9 studies were excluded due to publications reporting sample sizes as 'number of slices' rather than 'number of animals' or not making it clear if the sample size related to slices or animals. For the meta analysis to be suitable, the cut off n $\geq$ 10 studies is used (Harrer et al. 2021). Random effects meta analysis was carried out using data from studies assessing LTP at 30 minutes post stimulation in CA1, LTP at 60 minutes post stimulation in CA1 and I/O in CA1 (Table 7).

Outcome	Brain Pathway	N comparisons	N publications
I/O Maximum	CA1	18	11
LTP (30 mins post-stimulation)	CA1	23	15
LTP (60 mins post-stimulation)	CA1	21	13
LTP (90 mins post-stimulation)	CA1	1	1
LTD (30 mins post-stimulation)	CA1	2	2
LTD (60 mins post-stimulation)	CA1	2	2
PPF (ISI 50)	CA1	16	4

Table 7: *Table detailing what comparisons can be used for meta-analysis, outcomes highlighted in red were excluded as n $<$ 10.*

3.4.1 Long Term Potentiation at 30 minutes in the CA1

In modelling experiments measuring LTP 30 minutes post-stimulation, the pooled effect of transgenic modelling was -0.7055 (95% CI –1.1 to -0.3, n = 23 comparisons), meaning that APP animal models had reduced LTP compared to the wild-type controls. It was found that heterogeneity was high across these comparisons however (I^2 = 64%, Q= 61.36, Tau^2 = 0.5141). A forest plot of comparisons is shown in Figure 8.

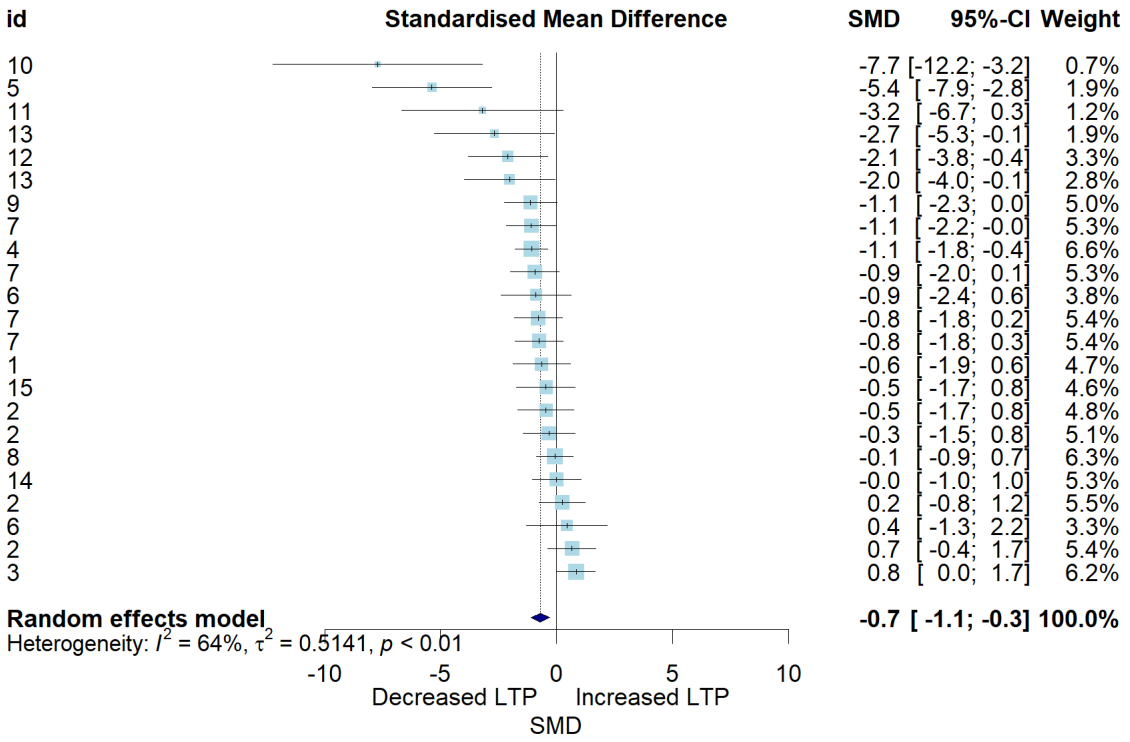


Figure 8: *Forest plot of random effects meta analysis of LTP after 30 minutes in the CA1 region. The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. The publications that related to the id numbers can be found in Appendix 5.*

3.4.2 Long Term Potentiation at 60 minutes in the CA1

In modelling experiments measuring LTP 60 minutes post-stimulation, the pooled effect of transgenic modelling was -0.9001 (95% CI -1.4412 to -0.3591, n = 21 comparisons), meaning that APP animal models showed even lower LTP compared to the wild-type controls at this point post stimulation compared with 30 minutes however heterogeneity was even higher (I^2 = 74.2%, Q= 77.61, Tau^2 = 1.0953). A forest plot of comparisons is shown in Figure 9.

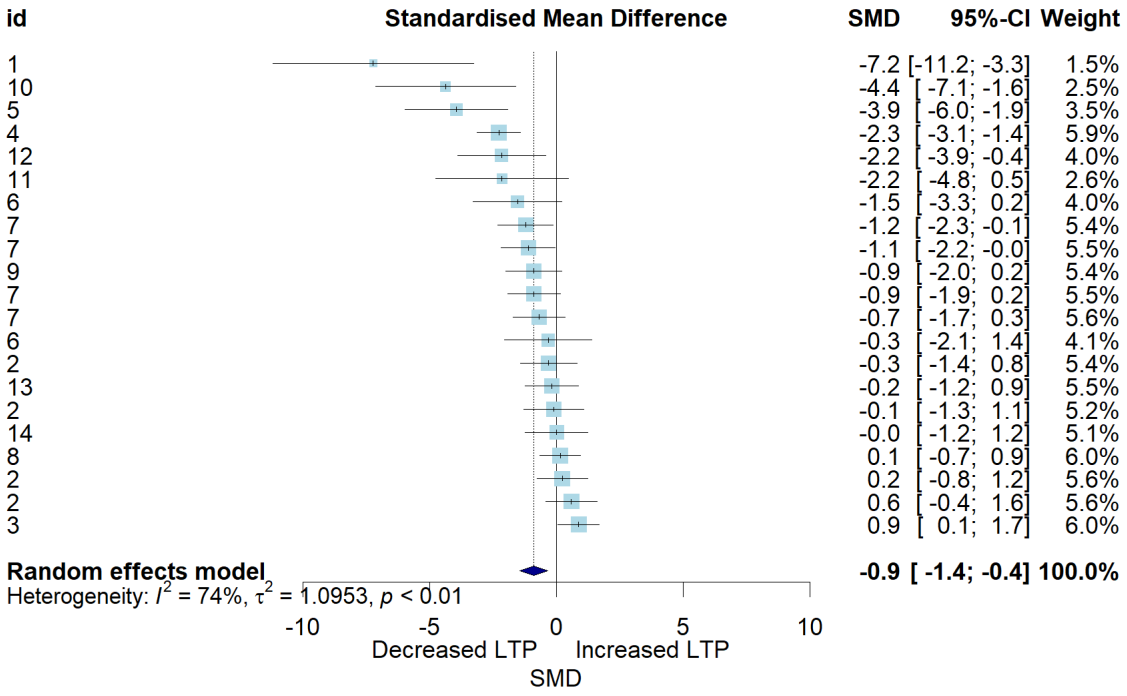


Figure 9: *Forest plot of random effects meta analysis of LTP after 60 minutes in the CA1 region. The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. The publications that related to the id numbers can be found in Appendix 5*

3.4.3 Input/ Output Transmission in the CA1

In modelling experiments measuring the maximum input/output (basal synaptic transmission) in the CA1 of hippocampal slices, the pooled effect of transgenic modelling was -0.2607 SMD (95% CI -0.5633 to 0.0419, n = 18 comparisons). Overall, transgenic animals had a slightly reduced input/output relationship versus wildtype controls and heterogeneity between studies was much lower (I^2 = 40.2%, Q = 28.43, Tau^2 = 0.1201). A forest plot of comparisons is shown in Figure 10.

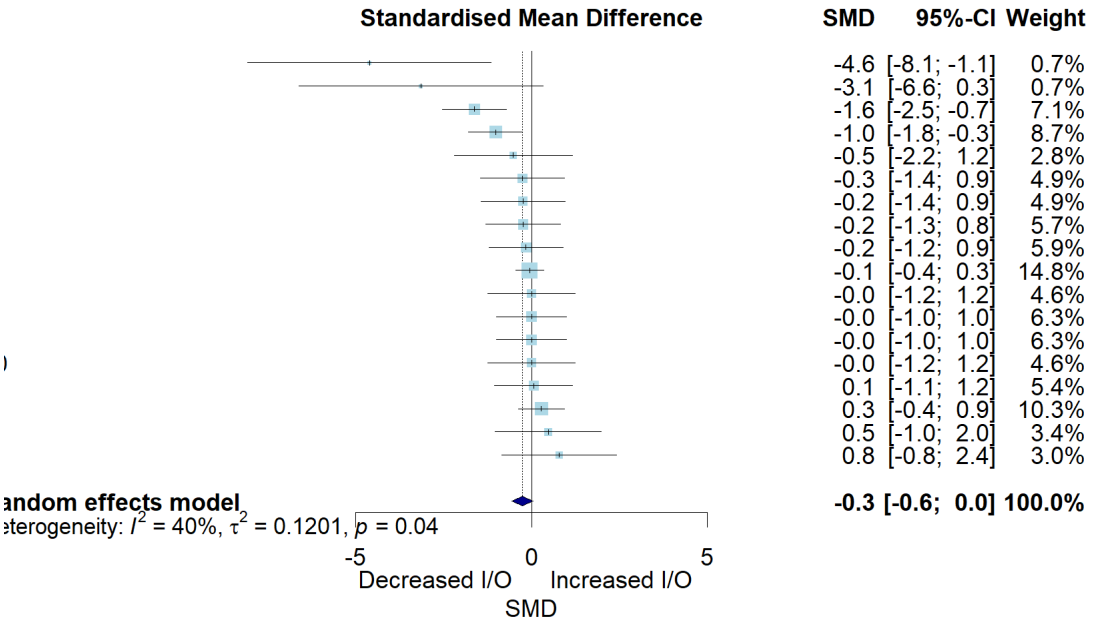


Figure 10: *Forest plot of random effects meta analysis of I/O in the CA1 region. The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. The publications that related to the id numbers can be found in Appendix 5*

3.5 Multi-Variable Meta Regression

The multi-variable meta regression to investigate sources of heterogenity was not possible due to lack of data, with none of the variables meeting the threshold of reporting across 25 experiments.

3.6 Meta Regression Analysis: Synaptic function and cognition

The regression planned to investigate the relationship between the Morris water maze outcomes and electrophysiology was not possible as less than 20% of studies (n=4) identified included MWM outcomes and of these, none made it clear if the same animals used for the behavioural test and the electrophysiology outcomes.

3.7 Publication Bias

As per the protocol (Appendix 1) the only data set with enough comparisons for publication bias analysis was LTP at 30 minutes post stimulation.

3.7.1 Long Term Potentation at 30 minutes in the CA1

A visual inspection of the funnel plot shown in Figure 11A indicates asymmetry, suggesting there could be publication bias present as it shows small study effects, meaning smaller studies showed a larger modelling effect than larger studies (Choi et al. 2018). However when Egger's test was performed it did not indicate funnel plot asymmetry for LTP outcomes at 30 minutes post stimulation (t = -0.92, df = 21 , p-value = 0.3677).

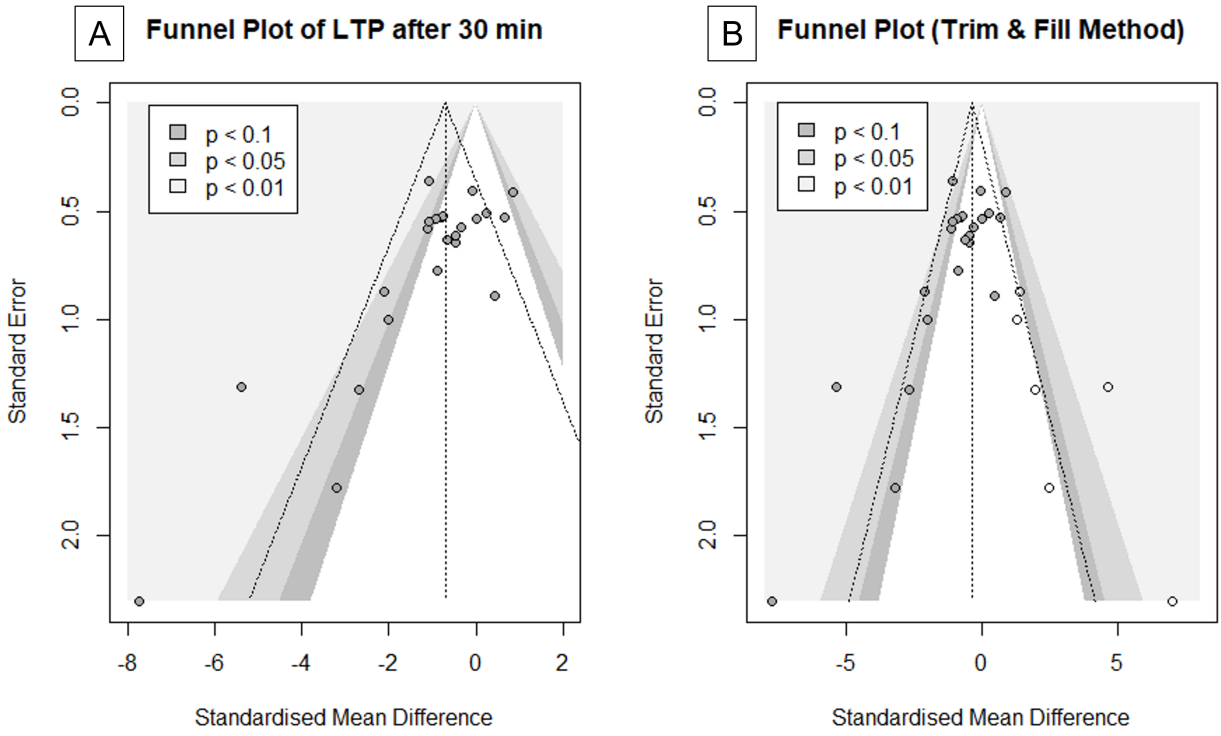


Figure 11: *Contour-Enhanced Funnel Plot showing effect sized of **A**; LTP after 30 min and **B**; LTP after 30 min with 6 theoretical missing studies added and shown as unfilled circles using the trim and fill method. Filled circles represent reported experiments.*

The trim and fill analysis results added an additional 6 theoretically missing studies in order to assess the magnitude of the effect of asymmetry on synaptic plasticity outcomes if present (Figure 11B). Interestingly this did decrease the estimate of synaptic plasticity deficits from the original meta-analysis shown in Section 3.4.1 by 46.84% to -0.3748 SMD (95% CI -0.9470 to 0.1975) which again may indicate the presence of publication bias.

3.7.2 Post Hoc Sensitivity Test

As investigations into the extent of publication bias showed mixed evidence of small study effects, a sensitivity test was carried out post hoc. The trim and fill method often does not produce reliable results when heterogeneity is high (Simonsohn et al. 2015), which is the case in this analysis (I^2 = 64% for LTP at 30 minutes). The sensitivity test involves identifying and removing extreme outliers from the data set using the techniques set out by Harrer et al. (2021) to reduce heterogeneity. A subsequent trim and

fill analysis is carried out with this new dataand compared to the original result. Three outliers were identified for LTP at 30 minutes (studies 10, 3 and 5) which reduced the heterogeneity significantly (I^2 = 29.1% , Q =26.79, Tau^2=0.1009). Trim and fill added 3 theoretically present studies and the result still gave a reduced estimate of synaptic plasticity deficits but only by 27.01% -0.5151 SMD (95% CI -0.8199 to -0.2104) compared with the original trim and fill which showed a reduced estimate by 46.84%.

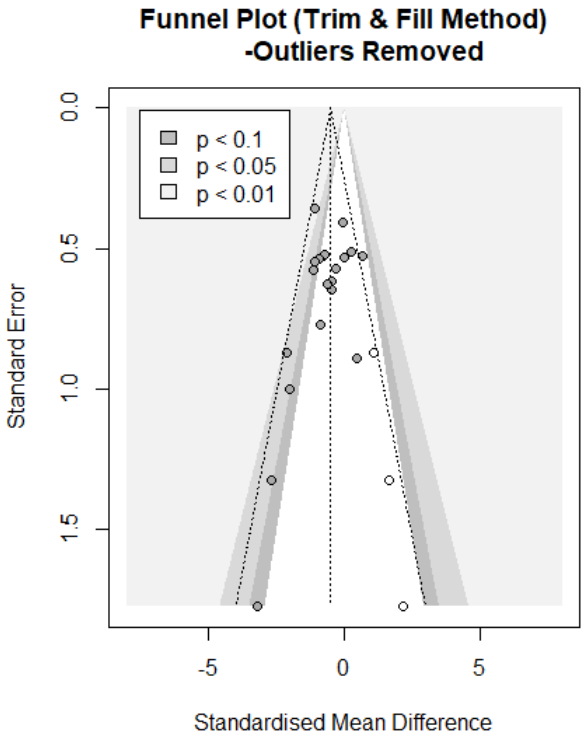


Figure 12: *Sensitivity Test: Trim and fill analysis with outliers removed.*

4 Discussion

This systematic review and meta-analysis assessed results across 24 publications. Some of the key issues highlighted by the results are discussed in this section.

4.1 Random Effects Meta-Analysis

The random effects meta-analysis (Section 3.4) indicates that the heterogeneity across experiments was high, especially in comparisons looking at LTP data ($I2$ = 74.2% in LTP at 60 mins post stimulation) with some studies reporting increased synaptic function in APP models compared to WT and others showing deficits. Despite this heterogeneity, the results showed that APP models of AD have reduced measurements of synaptic plasticity and I/O relationships (baseline synaptic strength) compared to wild type animals. Each of the comparisons showed a very strong pooled effect of transgenic modelling, with LTP at 60 minutes indicating the strongest pooled effect with -0.9001 SMD (95% CI -1.4412 to -0.3591).

4.2 Modelling of Alzheimer's Disease in APP models

Across the 15 different models investigated in the 24 studies included in this review, only 3 were investigated more than once; The Tg2576 (n = 5), J20 model (n = 4) and APP NL-G-F knock-in (n = 3) models. With thes models known to develop amyloid pathology at different stages of development ranging from 12 to 36 weeks (Mehla et al. 2019, Yoshiyama et al. 2015, Mucke et al. 2000, Hsiao et al. 1996), the younger ages preferred by researchers as shown in Figure 4 may have had an impact on synaptic plasticity deficits and account for some of the heterogeneity seen across publications. The insufficient levels of independent verification of results across most of the models, with 12 models only investigated once, may be reflective of the wider reproducibility crisis across the AD research field (Brownlee et al. 2021). It points towards the need for stronger effort to increase replication efforts so a better understanding of the effect of modeling on synaptic plasticity outcomes can be developed.

4.3 Psudoreplication and Sample Size

Over a third of experiments identified for this review (37.5%) did not appropriately report sample sizes for experiments. Psudoreplication would be introduced if sample size was taken to be the number of slices, as the non-independent nature of the data could lead to an inflated estimate of effect (Lazic et al. 2018). Unfortunately to avoid introducing psudoreplication into the analysis, any study that was unclear about sample size reporting or that reported number of slices was excluded, limiting the full utilisation of evidence available. This meant that the multi-variable meta-regressions set out in the protocol to determine statistically significant sources of heterogeneity were too under-powered to perform. On top of this, none of the publications 24 publications reported that they carried out a sample size calculation. The combination of these factors means the data must be interpreted with caution. Deficits in synaptic function in APP transgenic animal models may appear overstated and this is could be important for the analysis of the efficacy of any treatments developed.

4.4 Publication bias

After the initial investigation into publication bias in LTP outcomes 30 minutes post stimulation, the results of the funnel plot, Egger's Test and Trim and fill reported in Section 3.7 gave mixed evidence of small study effects. With the additional post hoc sensitivity test, the results indicated that the asymmetry seen in the original funnel plot was likely caused by factors other than publication bias, such as heterogeneity (I^2 = 64%) or small study effects (Harrer et al. 2021). The previous systematic review by Hair (2021) found evidence of publication bias in LTP outcomes 30 minutes post stimulation suggesting global effect sizes in that data set may have been overstated by as much as 53%. Though the current project only investigated one model type, the lack of publication bias present compared to the systematic review carried out pre-2018 may indicate that initiatives, such as preregistered reports are having a positive effect on the data available to researchers.

4.5 Limitations

There are a number of limitations to this study. All data included was screened and extracted by a single reviewer, which could mean up to 13% of relevant studies were missed at screening and extracting errors included in the analysis (Dunn et al. 2020). As shown in Sections 3.5 and 3.6, a significant portion of the analysis was not possible due to the limited number of papers and poor sample size reporting. Prioritising studies with a regex match of $>$1 also meant it was likely that around 5% of studies were missing from the data set (1-2 studies).

4.6 Future Directions

The main improvement for this study would be to widen the size of the data set for the period 2018-2023. This could be best achieved by increasing the scope of this study to include other transgenic model types including double APP/PSN1 or the triple APP/PSN1/MAP1 models as well as other models with mutations in tau. Once complete, an analysis could be carried out combining the enlarged dataset with one gathered by Hair (2021). This larger meta-analysis would have greater statistical power to determine the difference in synaptic plasticity / transmission outcomes for animal models of AD compared to WT. A larger data set my allow for causes of heterogeneity to be identified more easily which could inform better research design and reporting in the future.

5 Conclusions

The conclusions of the study are as follows:

- The results of the random effects meta-analysis showed that the impact of transgenic modeling of AD with mutations in the APP gene was to reduce synaptic plasticity outcomes. Each of the comparisons showed a very strong pooled effect compared to WT animals.
- The funnel plot, Egger's Test and Trim and fill results gave mixed evidence of small study effects. With the additional post hoc sensitivity test, the risk of bias reporting in the literature was determined to be low and have little impact on the effect sizes.
- There were many experimental design variables identified, many of which were poorly reported. The multi-variable meta-regression to investigate the influence of these variables was not possible due to lack of data caused by the limited number of studies and poor sample size reporting.

The recommendations made are:

- Widen the size of the data set for the period 2018-2023 to include other transgenic model types including double APP/PSN1 or the triple APP/PSN1/MAP1 models as well as models with mutations in tau.
- Conduct a meta-analysis to determine the difference in synaptic plasticity / transmission outcomes for animal models of AD compared to WT with this data and to identify sources of heterogeneity in the literature.

The greater understanding of baseline synaptic function in animal models of AD would be a highly useful tool for the drug development pipeline in order to understand the efficacy of disease altering treatments investigated in pre-clinical trials.
Wordcount: 4955

Appendix 1: Protocol

SYSTEMATIC REVIEW PROTOCOL FOR ANIMAL INTERVENTION STUDIES

FORMAT BY SYRCLE (WWW.SYRCLE.NL)
VERSION 2.0 (DECEMBER 2014)

Item #	Section/Subsection/Item	Description	Check for approval
	A. General		
1.	Title of the review	Systematic review and meta-analysis of synaptic plasticity in animal models of Alzheimer's disease: an update	
2.	Authors (names, affiliations, contributions)	Gabriella Lambert, Laurel Renton	
3.	Other contributors (names, affiliations, contributions)	Kaitlyn Hair – kaitlyn.hair@ed.ac.uk Dr Kaitlyn Hair – Centre for Clinical Brain Sciences, The University of Edinburgh	
4.	Contact person + e-mail address	Gabriella Lambert – lambert.ella@gmail.com	
5.	Funding sources/sponsors	CAMARADES	
6.	Conflicts of interest	None	
7.	Date and location of protocol registration		
8.	Registration number (if applicable)		
9.	Stage of review at time of registration	Search validation conducted using regex search terms and search of literature performed however screening for inclusion and exclusion has not yet started.	
	B. Objectives		
	Background		
10.	What is already known about this disease/model/intervention? Why is it important to do this review?	Alzheimer's Disease (AD) is a progressive neurodegenerative disorder that causes cognitive decline and severe memory impairment. AD pathology involves senile plaques made from a build-up of β Amyloid (Aβ) peptides, neurofibrillary tangles (NFTs) consisting of hyperphosphorylated tau protein and neuroinflammation, leading to brain atrophy and synapse loss (Shineman *et al.*, 2011). Transgenic modelling involves altering the genome of an animal to include a transgene or foreign sequence in order to study the gene function and model disease (Kitazawa *et al.*, 2012). Development of these animal models for studying AD is largely focused on the amyloid theory which hypothesises that the Aβ build-up in the brain is the primary driver of AD pathology (Hardy and Allsop, 1991). This has also been the focus when it comes to producing therapeutics. However emerging evidence suggests that there are irregularities in synapses early on in AD progression, and synaptic loss in the hippocampus is most highly correlated with cognitive impairment (Terry *et al.*,	

		1991). The idea that AD is a disease of synaptic failure suggests there may be therapeutic potential in targeting synaptic dysfunction. Synapse loss in AD is something that is widely studied using a measure of hippocampal synaptic plasticity in the form of long term potentiation (LTP) and depression (LTD) (Bliss and Collingridge, 1993). The loss of synaptic plasticity is hypothesised to be a precursor to synapse loss and is a widely accepted neuronal model responsible for learning and memory formation measured using electrophysiology (Egan *et al.*, 2016). To assess learning and memory directly from animal behaviour, many authors use the Morris water maze (MWM) paradigm. With the emerging role of synaptic dysfunction in cognitive impairment in AD, it important to establish whether synaptic dysfunction is linked to cognitive impairment in animal models for drug development purposes. Though synaptic plasticity has been described as altered in animal models of AD, across the current literature there are several discrepancies and conflicts surrounding this (Marchetti and Marie, 2011). Therefore the aim of the current systematic review is to collect and summarise all AD studies investigating in-vitro electrophysiology outcomes in combination with behavioural assessments using the Morris Water Maze paradigm, to assess the current research pertaining to synaptic plasticity in animal models to better understand the discrepancies seen across the literature and to get a better idea of this property for the development of therapeutics. The transgenic models being considered in these studies are the APP (amyloid precursor protein) model and the APP/PSEN1 (presenilin-1, a catalytic subunit of γ-secretase which cleaves APP) model (Elder *et al.*, 2010). It will cover all literature from 2018-2023 (all pre 2018 data has already been collected).	
	Research question		
11.	Specify the disease/health problem of interest	Alzheimer's disease	
12.	Specify the population/species studied	APP and combined APP/PSEN1 animal models	
13.	Specify the intervention/exposure	No intervention (evaluating modelling effect only)	
14.	Specify the control population	Wild type rodents (littermates or age-matched of same background strain)	
15.	Specify the outcome measures	- *In vitro* electrophysiology - field recordings of hippocampal slice long-term potentiation (LTP)	

		and depression (LTD) as well as synaptic transmission. - Neurobehavioural – spatial learning and memory assessed with Morris Water Maze (MWM) - Reporting quality – number of times bias-reducing measures are reported	
16.	State your research question (based on items 11-15)	- Which study design factors are transparently reported in studies with *in vitro* electrophysiology and MWM, and their impact on results? - Which variables increase risk of bias in *in vitro* electrophysiology and MWM experiments, how frequently do they appear in the literature, and how do they impact effect sizes? - Which aspects of model design influence external validity of transgenic AD animals in their recapitulation of synaptic transmission/plasticity, and how do these link to cognitive results? - How do electrophysiological experiments of transgenic AD animal models compare within each other and to human AD?	
	C. Methods		
	Search and study identification		
17.	Identify literature databases to search (*e.g.* Pubmed, Embase, Web of science)	☐MEDLINE via PubMed ☐Web of Science ☐SCOPUS ☐EMBASE ☒Other, namely: AD-SOLES database (https://camarades.shinyapps.io/ad-soles/) (contains all other databases) ☐Specific journal(s), namely:	
18.	Define electronic search strategies (*e.g.* use the step by step search guide[15] and animal search filters[20, 21])	When available, please add a supplementary file containing your search strategy: [insert file name]	
19.	Identify other sources for study identification	☐Reference lists of included studies ☐Books ☐Reference lists of relevant reviews ☐Conference proceedings, namely: ☐Contacting authors/ organisations, namely: ☐Other, namely:	
20.	Define search strategy for these other sources		
	Study selection		
21.	Define screening phases (*e.g.* pre-screening based on title/abstract, full text screening, both)	Searched AD-SOLES database to retrieve 2109 studies, random sample of n=200 papers screened to validate the regex terms. Determined that >1 full text regex matches identified 95% of included papers (sensitivity 95%, specificity 43.8%). 1330 papers in the set added to SyRF project and prioritised for screening while papers that don't meet this threshold of matches will be reviewed at a later date.	

22.	Specify (a) the number of reviewers per screening phase and (b) how discrepancies will be resolved	2 reviewers per screening phase Discrepancies resolved with a third expert party (Dr Kaitlyn Hair)	
	Define all inclusion and exclusion criteria based on:		
23.	Type of study (design)	Inclusion criteria: All primary experiments in transgenic AD animal models, including conference abstracts. Exclusion criteria: All studies without proper control cohorts and non-primary studies including commentaries, reviews, letters, and editorials.	
24.	Type of animals/population (*e.g.* age, gender, disease model)	Inclusion criteria: Primary experiments in transgenic animals of any sex, age, or species. Exclusion criteria: Primary experiments in non-transgenic or combined AD animal models, or studies in humans, *in vitro*, *ex vivo*, or *in silico*.	
25.	Type of intervention (*e.g.* dosage, timing, frequency)	Treatment data not being extracted however data from studies that compare a WT control and an AD animal model control (either both naïve or both been administered a vehicle or control intervention) will be included.	
26.	Outcome measures	Inclusion criteria: *In vitro* electrophysiological experiments assessing synaptic transmission/plasticity, with a focus on those assessing cognition with Morris Water Maze experiments. Exclusion criteria: Any study which does not involve electrophysiological experiments assessing synaptic transmission/plasticity	
27.	Language restrictions	Inclusion criteria: All languages Exclusion criteria: None	
28.	Publication date restrictions	Inclusion criteria: All publication dates Exclusion criteria: None	
29.	Other	Inclusion criteria: Exclusion criteria:	
30.	Sort and prioritize your exclusion criteria per selection phase	Selection phase: full text screening* 1. all previously stated inclusion/exclusion criteria will be applied in this phase *most studies do not report electrophysiology in title/abstract so full text screening and data extraction are being carried out in one phase.	
	Study characteristics to be extracted (for assessment of external validity, reporting quality)		
31.	Study ID (*e.g.* authors, year)	- Title - Primary author - Corresponding authors - DOI - Year	

		- Journal name - Country of origin of corresponding author	
32.	Study design characteristics (*e.g.* experimental groups, number of animals)	Animal husbandry - Light cycle of animal facilities - Number of animals per cage - Environmental enrichment reported	
33.	Animal model characteristics (*e.g.* species, gender, disease induction)	- Control procedure (WT littermates or age-matched from same background strain) - Animal species - Animal background strain - Source of animals - Sex of animals - Age of animals	
34.	Intervention characteristics (*e.g.* intervention, timing, duration)	Treatment effect not assessed	
35.	Outcome measures	In vitro electrophysiological field potential recordings: - Outcome measure average type - Outcome measure error type (SD or SEM) - Brain pathway recorded (e.g. Dentate gyrus DG) - Outcome measure category (e.g. LTP/LTD) - Outcome measure units - Outcome measure detail (e.g. % Normalised fEPSP slope) - Time taken (minutes) of baseline recording - Blockers used while investigating synaptic plasticity/transmission - Time of day animals were sacrificed for ephys - Anaesthesia used prior to decapitation - Use of kynurenic acid during brain dissection - Direction of brain slice - Duration (minutes) for brain slice recovery after slicing - Temperature slices left to recover at - Percentage of maximal response from input/output used for baseline recordings - Type of recording chamber used - Calcium concentration of slicing solution - Magnesium concentration of slicing solution - Calcium concentration of recording solution - Magnesium concentration of recording solution - LTP/LTD induction protocols: o Type of stimulation used to induce LTP o Number of stimulation trains o Inter-stimulation interval o Number of bursts in 1 stimulation train o Frequency (Hz) of bursts in 1 stimulation train o Inter-burst interval in 1 stimulation train o Inter-burst interval units o Number of pulses in 1 burst o Frequency (Hz) of pulses in 1 burst	

		○ Inter-pulse interval in 1 burst ○ Inter-pulse interval units ○ Percentage (%) of maximal response used for induction ○ Total number of stimulations Morris Water Maze paradigm: - Type of outcome (acquisition vs. probe) - Number of training days - Number of trials per day - Whether a visible platform test was used - Pool temperature - Acquisition phase outcome measure unit (e.g. escape latency (s) / training days)	
36.	Other (*e.g.* drop-outs)		
	Assessment risk of bias (internal validity) or study quality		
37.	Specify (a) the number of reviewers assessing the risk of bias/study quality in each study and (b) how discrepancies will be resolved	2 reviewers will assess the risk of bias/ study quality in each study. Discrepancies resolved with a third expert party (Dr Kaitlyn Hair)	
38.	Define criteria to assess (a) the internal validity of included studies (*e.g.* selection, performance, detection and attrition bias) and/or (b) other study quality measures (*e.g.* reporting quality, power)	☐By use of SYRCLE's Risk of Bias tool[4] ☐By use of SYRCLE's Risk of Bias tool, adapted as follows: ☐By use of CAMARADES' study quality checklist, e.g [22] ~~☐~~By use of CAMARADES' study quality checklist, adapted as follows: - Reporting of random allocation of animals to experimental groups - Reporting of blinding during outcome assessment - Reporting of a sample size calculation - Reporting of a study protocol dated before the study began - Reporting of potential conflicts of interest statement - Reporting of rationale for data point inclusion/exclusion Reporting whether experiments were approved by an animal welfare committee ☐Other criteria, namely:	
	Collection of outcome data		
39.	For each outcome measure, define the type of data to be extracted (*e.g.* continuous/dichotomous, unit of measurement)	Electrophysiological outcomes: - Continuous for all numerical variables (e.g. # stimulation trains) - Dichotomous for categorical variables (e.g. type of stimulation used to induce LTP) Behavioural (MWM) outcomes: - Continuous for all numerical variables (e.g. # training days) - Dichotomous for categorical variables (e.g. acquisition vs. probe)	

40.	Methods for data extraction/retrieval (*e.g.* first extraction from graphs using a digital screen ruler, then contacting authors)	Extract whether sample size (N) refers to number of hippocampal slices (which would be a source of pseudoreplication) or individual animals (the experimental unit). We then extract numerical data (mean, SD or SEM) from the full texts of included papers using WebPlotDigitizer.	
41.	Specify (a) the number of reviewers extracting data and (b) how discrepancies will be resolved	Two reviewers will extract data from each included publication and resolve discrepancies by discussion,	
	Data analysis/synthesis		
42.	Specify (per outcome measure) how you are planning to combine/compare the data (*e.g.* descriptive summary, meta-analysis)	Findings from SR will be collated in a descriptive summary, paired with a meta-analysis of numerical data	
43.	Specify (per outcome measure) how it will be decided whether a meta-analysis will be performed	All numerical data will be included in a meta-analysis, with more weight given to studies with higher precision	
	If a meta-analysis seems feasible/sensible, specify (for each outcome measure):		
44.	The effect measure to be used (*e.g.* mean difference, standardized mean difference, risk ratio, odds ratio)	Standardised mean difference	
45.	The statistical model of analysis (*e.g.* random or fixed effects model)	Multivariable meta-regression, if sufficient sample size ($n \geq 10$). Alternatively, each characteristic will be investigated using a univariate approach.	
46.	The statistical methods to assess heterogeneity (*e.g.* I^2, Q)	Cochran's, Tau^2, residual I^2	
47.	Which study characteristics will be examined as potential source of heterogeneity (subgroup analysis)	Study quality measures, model type, animal population characteristics (age, sex), electrophysiology protocol characteristics.	
48.	Any sensitivity analyses you propose to perform	Sensitivity analysis in R	
49.	Other details meta-analysis (*e.g.* correction for multiple testing, correction for multiple use of control group)	Where a control wild type control serves more than one modelling group, the number of animals in the control group will be corrected (Vesterinen *et al.*, 2016). Where several measurements of the same underlying biological outcome are reported from the same cohort of animals at the same time point, we will nest these comparisons into a single fixed effect (Vesterinen *et al.*, 2016).	
50.	The method for assessment of publication bias	Funnel plot, egger's and trim and fill will be used to assess publication bias for each outcome (requires at least 25 comparisons)	
Final approval by (names, affiliations):		Date:	

References

1. Bliss, T. V. P. and G. L. Collingridge (1993). A synaptic model of memory: long-term potentiation in the hippocampus. Nature (London) 361(6407) 31-39.
2. Egan, K. J., et al. (2016). From a mouse: systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. Evidence-based preclinical medicine 3(1) 12-23.
3. Elder, G. A., M. A. Gama Sosa and R. De Gasperi (2010). Transgenic Mouse Models of Alzheimer's Disease: G. A. ELDER ET AL.: TRANSGENIC MODELS AND ALZHEIMER'S DISEASE. The Mount Sinai journal of medicine 77(1) 69-81.
4. Hardy, J. and D. Allsop (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. Trends in Pharmacological Sciences 12(10) 383-388.
5. Kitazawa, M., R. Medeiros and F. M. Laferla (2012). Transgenic mouse models of Alzheimer disease: developing a better model as a tool for therapeutic interventions.
6. Marchetti, C. and H. Marie (2011). Hippocampal synaptic plasticity in Alzheimer's disease: what have we learned so far from transgenic models? Reviews in the neurosciences 22(4) 373-402.
7. Shineman, D. W., et al. (2011). Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. Alzheimers Res Ther 3(5) 28.
8. Terry, R. D., et al. (1991). Physical basis of cognitive alterations in alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. Annals of neurology 30(4) 572-580.
9. Vesterinen, H. M., et al. (2016). Meta-analysis of data from animal studies: A practical guide (vol 221, pg 92, 2014). Journal of Neuroscience Methods 259 156-156.

Appendix 2: Regex Expression

```
[Hh]ippocampal\.slice\.|[tT]ransverse\.vibratome\.sections|[tT]ransverse\.
    slice\.|[Ee]lectrophysi\.|[fF]ield\.(excitatory|potential|postynaptic)|
    fEPSP\.|[Ff]ield\ EPSP\.
```

Appendix 3: Pilot Study

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 in 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: Ss]lice.[pP]reparation|[Hh]ippocampal.slice.|[tT]ransverse.vibratome.sections|[tT]ransverse.slice.|[Ee]lectrophysi.| [fF]ield.(excitatory|potential|postynaptic)|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 Reveiw 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 reveiwers.

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 Reveiw 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 radom 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 appeopriatly.

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

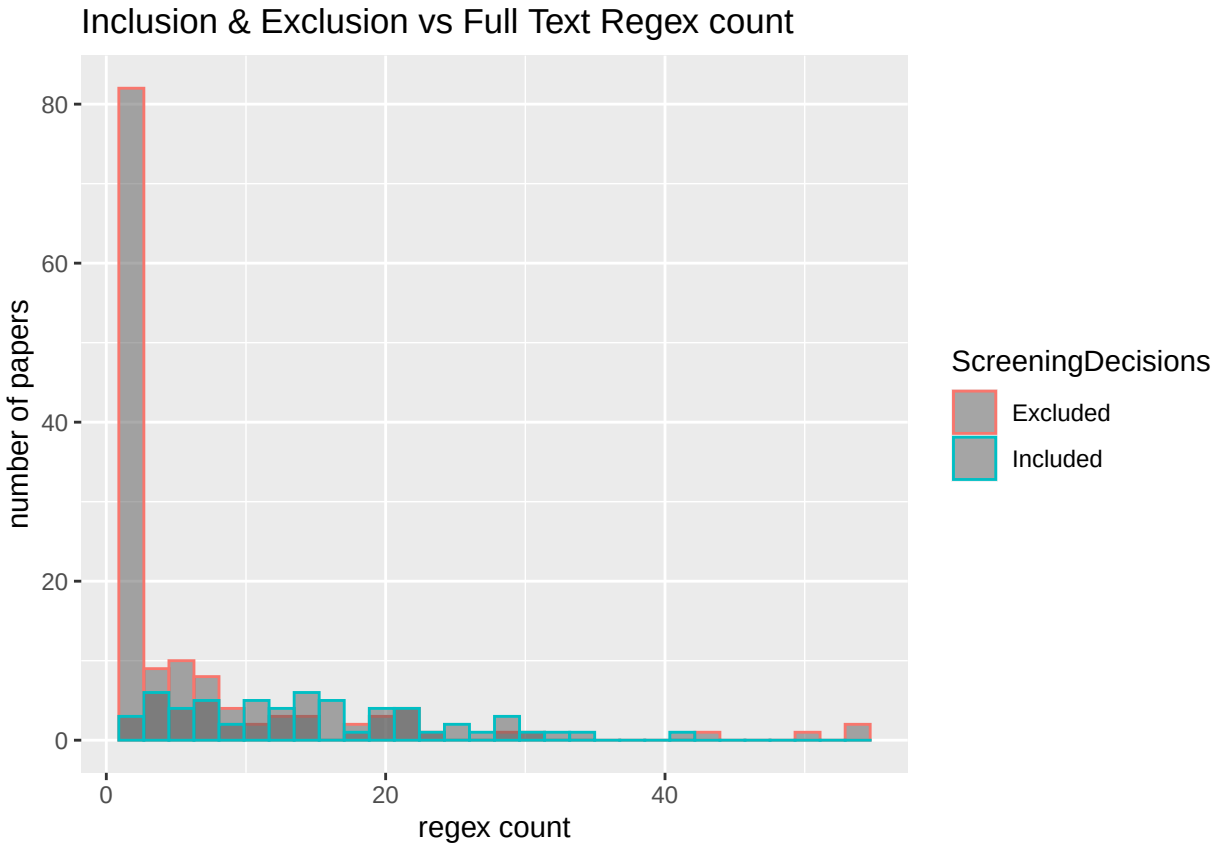
Manual screening decisions of the for inclusion and exclusion was uploaded and plotted against regex count. Regex count refers to the number of times the software picked up the "regualr expresions" serched for in the full text of the paper.

```
# 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;

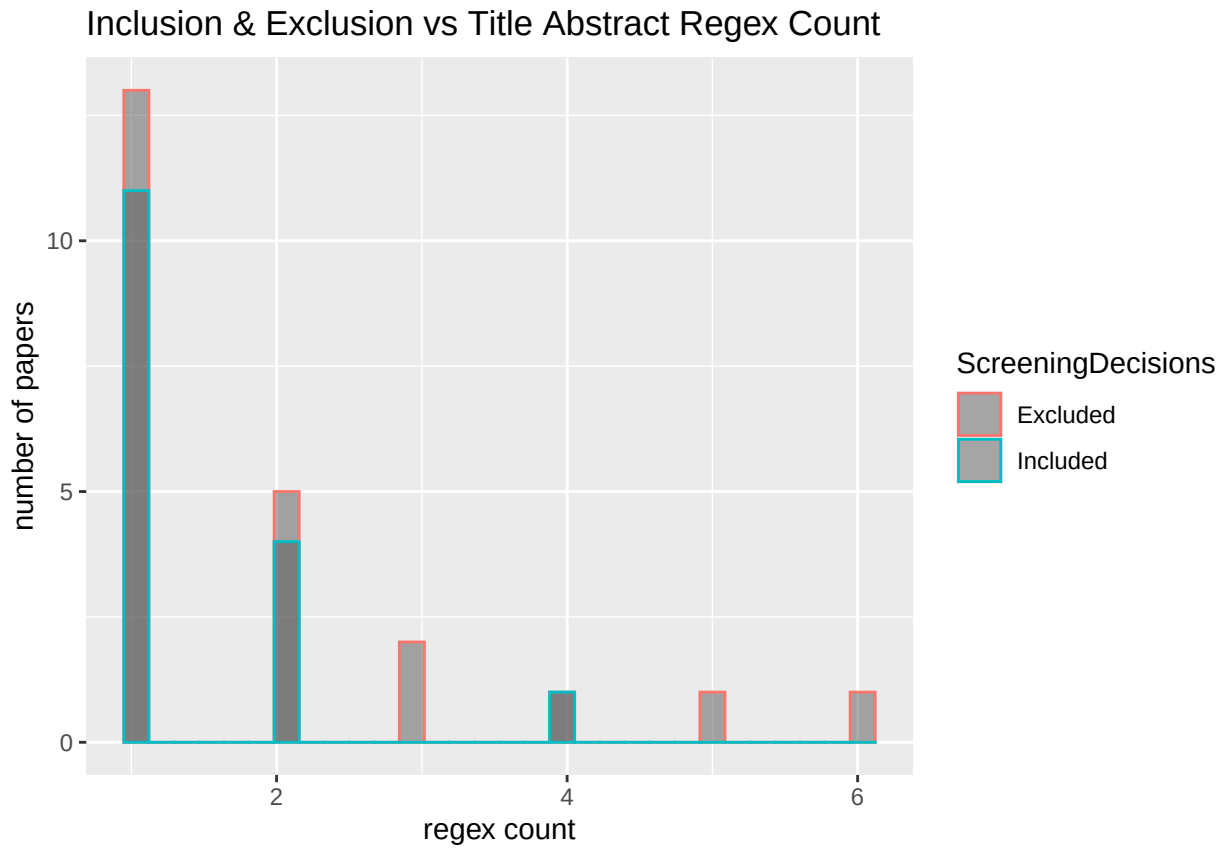
- true positive = a study that was included at manual screening and correctly included with a regex count cut-off at >1
- true negative = a study that should be excluded at manual screening is excluded correctly with the regex count cut-off of >1
- false positives = a study that should be excluded at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



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 used 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 used 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.

Appendix 4: Publications used in Meta-Analysis

Study ID	Reference
1	Wang, M., Ramasamy, V. S., Samidurai, M., & Jo, J. (2019). Acute restraint stress reverses impaired LTP in the hippocampal CA1 region in mouse models of Alzheimer's disease. Scientific Reports, 9(1), 10955. doi: 10.1038/s41598-019-47452-6.
2	Benitez, D. P., Jiang, S., Wood, J., et al. (2021). Knock-in models related to Alzheimer's disease: synaptic transmission, plaques and the role of microglia. Molecular Neurodegeneration, 16(1), 47. doi: 10.1186/s13024-021-00457-0.
3	Kreis, A., Desloovere, J., Suelves, N., et al. (2021) Overexpression of wild-type human amyloid precursor protein alters GABAergic transmission. Sci Rep, 11(1), 17600. https://doi.org/10.1038/s41598-021-97144-3
4	Corsetti, V. et al. (2020) "Passive immunotherapy for N-truncated tau ameliorates the cognitive deficits in two Mouse alzheimer's disease models," Brain Communications, 2(1). doi: https://doi.org/10.1093/braincomms/fcaa039.
5	Cisternas, P. et al. (2019) "Presymptomatic treatment with andrographolide improves brain metabolic markers and cognitive behavior in a model of early-onset alzheimer's disease," Frontiers in Cellular Neuroscience, 13. doi: https://doi.org/10.3389/fncel.2019.00295.
6	Snow, W.M. et al. (2019) "Strain differences in hippocampal synaptic dysfunction in the TGCRND8 mouse model of alzheimer's disease: Implications for improving translational capacity," Molecular and Cellular Neuroscience, 94, pp. 11–22. Available at: https://doi.org/10.1016/j.mcn.2018.10.005.
7	Maki, T. et al. (2022) "App knock-in mice produce E22p-Aβ exhibiting an alzheimer's disease-like phenotype with dysregulation of hypoxia-inducible factor expression," International Journal of Molecular Sciences, 23(21), p. 13259. doi: https://doi.org/10.3390/ijms232113259.
8	Sarmi Sri, Chrysia-Maria Pegasiou, Chantal Abbigail Cave, Hough, K., Wood, N.J., Gomez-Nicola, D., Deinhardt, K., Bannerman, D.M., V. Hugh Perry and Vargas-Caballero, M. (2019). Emergence of synaptic and cognitive impairment in a mature-onset APP mouse model of Alzheimer's disease. Acta neuropathologica communications, 7(1). doi:https://doi.org/10.1186/s40478-019-0670-1.
9	Medawar, E., Benway, T.A., Liu, W., Hanan, T.A., Haslehurst, P., James, O.T., Yap, K., Muessig, L., Moroni, F., Nahaboo Solim, M.A., Baidildinova, G., Wang, R., Richardson, J.C., Cacucci, F., Salih, D.A., Cummings, D.M. and Edwards, F.A. (2019). Effects of rising amyloidβ levels on hippocampal synaptic transmission, microglial response and cognition in APPSwe/PSEN1M146V transgenic mice. EBioMedicine, [online] 39, pp.422–435. doi:https://doi.org/10.1016/j.ebiom.2018.12.006.

10	Vijay Sankar Ramasamy, Manikandan Samidurai, Hyung Ho Park, Wang, M., Park, R.-Y., Seon Oh Yu, Hee Gyung Kang, Hong, S., Won Suk Choi, Lee, Y.-Y., Hyung Sik Kim and Jo, J. (2020). Avenanthramide-C Restores Impaired Plasticity and Cognition in Alzheimer's Disease Model Mice. Molecular Neurobiology, 57(1), pp.315–330. doi:https://doi.org/10.1007/s12035-019-01707-5.
11	Du, F.K., Yu, Q., Shi Fang Yan, Zhang, Z., Jhansi Rani Vangavaragu, Chen, D., Shi Fang Yan and Shirley ShiDu Yan (2021). Gain of PITRM1 peptidase in cortical neurons affords protection of mitochondrial and synaptic function in an advanced age mouse model of Alzheimer's disease. Aging Cell, 20(5). doi:https://doi.org/10.1111/acel.13368.
12	Liu, Y., Cheng, A., Li, Y.-J., Yang, Y., Kishimoto, Y., Zhang, S., Wang, Y., Wan, R., Raefsky, S.M., Lu, D., Saito, T., Saido, T., Zhu, J., Wu, L.-J. and Mattson, M.P. (2019). SIRT3 mediates hippocampal synaptic adaptations to intermittent fasting and ameliorates deficits in APP mutant mice. Nature Communications, [online] 10(1). doi:https://doi.org/10.1038/s41467-019-09897-1.
13	Rutigliano, G., Stazi, M., Arancio, O., Watterson, D.M. and Origlia, N. (2018). An isoform-selective p38α mitogen-activated protein kinase inhibitor rescues early entorhinal cortex dysfunctions in a mouse model of Alzheimer's disease. Neurobiology of Aging, 70, pp.86–91. doi:https://doi.org/10.1016/j.neurobiolaging.2018.06.006.
14	Rancillac, A., Geoffroy, H. and Rossier, J. (2012). Impaired Neurovascular Coupling in the APPxPS1 Mouse Model of Alzheimer's Disease. Current Alzheimer Research, 9(10), pp.1221–1230. doi:https://doi.org/10.2174/156720512804142859.
15	Li, S., Jin, M., Liu, L., Dang, Y., Ostaszewski, B.L. and Selkoe, D.J. (2018). Decoding the synaptic dysfunction of bioactive human AD brain soluble Aβ to inspire novel therapeutic avenues for Alzheimer's disease. Acta Neuropathologica Communications, 6(1). doi:https://doi.org/10.1186/s40478-018-0626-x.
16	Sarmi Sri, Chrysia-Maria Pegasiou, Chantal Abbigail Cave, Hough, K., Wood, N.J., Gomez-Nicola, D., Deinhardt, K., Bannerman, D.M., V. Hugh Perry and Vargas-Caballero, M. (2019). Emergence of synaptic and cognitive impairment in a mature-onset APP mouse model of Alzheimer's disease. Acta neuropathologica communications, 7(1). doi:https://doi.org/10.1186/s40478-019-0670-1.
17	Moriguchi, S., Inagaki, R., Saito, T., Saido, T.C. and Fukunaga, K. (2022). Propolis Promotes Memantine-Dependent Rescue of Cognitive Deficits in APP-KI Mice. Molecular Neurobiology, 59(7), pp.4630–4646. doi:https://doi.org/10.1007/s12035-022-02876-6.
18	Baglietto-Vargas, D., Forner, S., Cai, L., Martini, A.C., Trujillo-Estrada, L., Swarup, V., Nguyen, M.M.T., Do Huynh, K., Javonillo, D.I., Tran, K.M., Phan, J., Jiang, S., Kramár, E.A., Nuñez-Diaz, C., Balderrama-Gutierrez, G., Garcia, F., Childs, J., Rodriguez-Ortiz, C.J., Garcia-Leon, J.A. and Kitazawa, M. (2021). Generation of a humanized Aβ expressing mouse demonstrating aspects of Alzheimer's disease-like pathology. Nature Communications, 12(1). doi:https://doi.org/10.1038/s41467-021-22624-z.

19	Altered inhibitory function in hippocampal CA2 contributes in social memory deficits in Alzheimer's mouse model.
20	Rancillac, A., Geoffroy, H. and Rossier, J. (2012). Impaired Neurovascular Coupling in the APPxPS1 Mouse Model of Alzheimer's Disease. Current Alzheimer Research, 9(10), pp.1221–1230. doi:https://doi.org/10.2174/156720512804142859.
21	Confettura, A.D., Cuboni, E., Ammar, M.R., Jia, S., Gomes, G.M., Yuanxiang, P., Raman, R., Li, T., Grochowska, K.M., Ahrends, R., Karpova, A., Dityatev, A. and Kreutz, M.R. (2022). Neddylation-dependent protein degradation is a nexus between synaptic insulin resistance, neuroinflammation and Alzheimer's disease. Translational Neurodegeneration, 11(1). doi:https://doi.org/10.1186/s40035-021-00277-8.
22	Beckelman, B.C., Yang, W., Kasica, N.P., Zimmermann, H.R., Zhou, X., Keene, C.D., Ryazanov, A.G. and Ma, T. (2019). Genetic reduction of eEF2 kinase alleviates pathophysiology in Alzheimer's disease model mice. Journal of Clinical Investigation, 129(2), pp.820–833. doi:https://doi.org/10.1172/jci122954.
23	Spoleti, E., Paraskevi Krashia, Livia La Barbera, Nobili, A., Carmen Alina Lupascu, Giacalone, E., Keller, F., Migliore, M., Rossi, M. and Marcello D'Amelio (2022). Early derailment of firing properties in CA1 pyramidal cells of the ventral hippocampus in an Alzheimer's disease mouse model. Experimental Neurology, 350, pp.113969–113969. doi:https://doi.org/10.1016/j.expneurol.2021.113969.
24	Latif-Hernandez, A., Sabanov, V., Ahmed, T., Craessaerts, K., Saito, T., Saido, T. and Balschun, D. (2020). The two faces of synaptic failure in AppNL-G-F knock-in mice. Alzheimer's Research & Therapy, 12(1). doi:https://doi.org/10.1186/s13195-020-00667-6.

Appendix 5: Code for Figures

Figures

Data

##Load Data

```
study_data <- read_csv("study_data2.csv")
diseasemodel_data <- read_csv("diseasemodel_data2.csv")
outcome_data <- read_csv("outcome_data2.csv")
```

My Data

```
study_data <- study_data %>%
    filter(Year >= 2018 & Year <= 2023, SystematicSearchName ==
"2020-2023 Update", InvestigatorName == "Laurel Renton",
StudyId != "6d537dfe-3d19-4dbf-b9af-bf9fad52bb4b")

outcome_data <- outcome_data %>%
    filter(Year >= 2018 & Year <= 2023, SystematicSearchName ==
        "2020-2023 Update", InvestigatorName == "Laurel Renton",
        StudyId != "6d537dfe-3d19-4dbf-b9af-bf9fad52bb4b")

diseasemodel_data <- diseasemodel_data %>%
    filter(Year >= 2018 & Year <= 2023, SystematicSearchName ==
        "2020-2023 Update", InvestigatorName == "Laurel Renton",
        StudyId != "6d537dfe-3d19-4dbf-b9af-bf9fad52bb4b")
```

Reporting Data

```
full_data_figures <- read.csv("full_data.csv") %>%
    filter(AnnotatorIdStr == "00374321-95a7-4a2a-a12c-af4c0f9e599b",
        StudyId != "6d537dfe-3d19-4dbf-b9af-bf9fad52bb4b")
```

Number of Studies Included

```
no_of_studies <- study_data %>%
    filter(QuestionId == "335bd414-ce8b-4127-a4ff-b02ec66a3b81",
        Answer == "Yes")
```

24 studies included after the screening process.

Transgenic APP Models of AD

Data Wrangling

```
diseasemodel_comments <- read_csv("comments.csv") %>%
    filter(InvestigatorName == "Laurel Renton") %>%
    select(StudyId, Answer, Comments)
# adds comments for the models that weren't listed in the
# SyRF project


model_data <- diseasemodel_data %>%
    left_join(diseasemodel_comments) %>%
    filter(QuestionId == "8255f1cf-1db1-4756-b77c-003df2860ed9") %>%
    select(StudyId, Answer, Comments) %>%
    unique() %>%
    mutate(Answer = if_else(!is.na(Comments), Comments, Answer),
        Answer = str_replace_all(Answer, "AppNL-G-F", "APP NL-G-F Knock-in"),
        Answer = str_replace_all(Answer, "NL-P-F", "APP NL-P-F Knock-in"),
        Count = 1) %>%
    select(Answer, Count) %>%
    group_by(Answer) %>%
    summarise(Count = sum(Count)) %>%
    filter(Answer != "PDGF-APP(WT) (line I5)")
```

Plotting Donut Plot

```
# Donut plot
hsize <- 4

color_palette <- c("#D1E7F6", "#75B6E5", "#77CEEF", "#1CADE4",
    "#2683C6", "#7AE0E5", "#42BA97", "#146266", "#A1C8C5", "#D3F5F6",
    "#BECAD4", "#597287", "#0E5772", "#134263", "#6E9281", "#59A79E")


donut_plot <- model_data %>%
    mutate(x = hsize) %>%
    ggplot(aes(x = hsize, y = Count, fill = Answer)) + geom_col(color = "black") +
    geom_text(aes(label = Count), position = position_stack(vjust = 0.5),
        size = 10, color = "white") + coord_polar(theta = "y") +
    scale_fill_manual(values = color_palette) + xlim(c(0.2, hsize +
    0.5)) + theme(panel.background = element_rect(fill = "white"),
    panel.grid = element_blank(), axis.title = element_blank(),
    axis.ticks = element_blank(), axis.text = element_blank(),
    legend.text = element_text(size = 12)) + guides(fill = guide_legend(title = "AD APP Model"))
donut_plot
```

Study Quality Protocols

```
quality_data <- full_data_figures %>%
select(StudyId, `Sample Size` = "Is.a.sample.size.calculation.reported.", Blinding =
"Does.the.paper.report.that.experimenters.were.blind.to.experimental.groups.during.outcome.
`Welfare Comitee Approval` =
"Does.the.paper.report.that.experiments.were.approved.by.an.animal.welfare.committee"
`Conflict of Interest` = "Is.there.a.conflicts.of.interest.statement.",
`Exclusion Criteria` = "Are.the.reasons.for.the.exclusion.of.data.points.reported.") %>%
unique() %>%
select(-StudyId) %>%
pivot_longer(cols = everything(), names_to = "study_quality_items",
values_to = "Value") %>%
group_by(study_quality_items, Value) %>%
summarize(Count = n()) %>%
pivot_wider(names_from = Value, values_from = Count) %>%
mutate(overall = 23, percentage_of_studies = `TRUE`/overall *
100, across(where(is.numeric), ~round(., 2)))

quality_data_plot <- quality_data %>%
ggplot(aes(x = reorder(study_quality_items, -percentage_of_studies),
y = percentage_of_studies)) + geom_bar(stat = "identity",
fill = "#174E77") + geom_text(aes(label = percentage_of_studies),
position = position_dodge(width = 0.9), hjust = 1.25, color = "white",
size = 3.5) + theme(plot.title = element_text(hjust = 0.5)) +
ggtitle("Studies (%) Reporting Different Study Quality Items") +
labs(x = "Study Quality Items", y = "Included (%)") + coord_flip() +
scale_y_continuous(labels = abs, limits = c(0, 100))

quality_data_plot
```

Electrophysiology reporting quality

```
## Applies to all experiments
ephys_data_1 <- full_data_figures %>%
filter(Outcome.category == "Electrophysiology") %>%
select(StudyId, Outcome = "OutcomeLabel",
 Age = "What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed."
`Brain Pathway` = "Brain.pathway.recorded",
`Type of Recording Chamber` = "What.type.of.recording.chamber.was.used"unique() %>%
select(-c(StudyId, Outcome, Age)) %>%
mutate(`Brain Pathway` = "Reported", `Type of Recording Chamber` = case_when(`Type of Recording Chamber`
"Not reported" ~ "NotReported", TRUE ~ "Reported")) %>%
pivot_longer(cols = c(`Brain Pathway`, `Type of Recording Chamber`),
names_to = "variable") %>%
group_by(variable, value) %>%
summarise(count = n()) %>%
pivot_wider(names_from = value, values_from = count) %>%
na_replace(0) %>%
mutate(Overall = Reported + NotReported)
```

```
## Applies to only experiments looking at LTP and LTD
ephys_data_2 <- full_data_figures %>%
    filter(Outcome.category == "Electrophysiology", OutcomeLabel %in%
        c("LTP", "LTD")) %>%
    select(StudyId, Outcome = "OutcomeLabel", `Stimulation Type (LTP)` = "What.type.of.stimulation.was.u
        `Stimulation Type (LTD)` = "What.type.of.stimulation.was.used.to.induce.LTD.",
        Age = "What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed.",
        `Total Stimulations` = "Total.number.of.stimulations",
        `% of Maximal Response of Baseline Recording` = "Ephys..percentage.....of.maximal.response.from
    unique() %>%
    select(-StudyId, -Age) %>%
    unite("Stimulation Type", c("Stimulation Type (LTP)", "Stimulation Type (LTD)")) %>%
    mutate(`LTP/LTD Stimulation Type**` = case_when(`Stimulation Type` ==
        c("Other (please leave a comment)_", "_Other (please leave a comment)") ~
        "NotReported", TRUE ~ "Reported"), `Total Stimulations**` = case_when(`Total Stimulations` ==
        "NR" ~ "NotReported", TRUE ~ "Reported"), `% of Maximal Response of Baseline Recording**` = case
        "NR" ~ "NotReported", TRUE ~ "Reported")) %>%
    select(-c(Outcome, `Stimulation Type`)) %>%
    pivot_longer(cols = c(`LTP/LTD Stimulation Type**`, `Total Stimulations**`,
        `% of Maximal Response of Baseline Recording**`), names_to = "variable") %>%
    group_by(variable, value) %>%
    summarise(count = n()) %>%
    pivot_wider(names_from = value, values_from = count) %>%
    mutate(Overall = Reported + NotReported)

ephys_data <- ephys_data_1 %>%
    full_join(ephys_data_2) %>%
    mutate(percentage = Reported/Overall, across(where(is.numeric),
        ~round(., 2))) %>%
    select(-c(NotReported, Reported, Overall))


ephys_table <- ephys_data %>%
    gt() %>%
    fmt_percent(columns = percentage, decimals = 1) %>%
    tab_header(title = md("**Reporting Quality of Electrophysiology Measures**")) %>%
    cols_label(percentage = "Experiments Reported*") %>%
    tab_source_note(source_note = "* 58 experiments across 24 studies assessing synaptic plasticity/ tra
    tab_source_note(source_note = "** across 28 experiments assessing LTP/LTD") %>%
    tab_style(locations = cells_column_labels(columns = everything()),
        style = list(cell_borders(sides = "bottom", weight = px(3)),
            cell_text(weight = "bold"))) %>%
    cols_width(percentage ~ "70%")

ephys_table
```

Animal Husbandary

```
animal_data <- full_data_figures %>%
    select(StudyId, lightcycle = "Animal.husbandry..light.cycle..h..of.animal.facilities",
        animalspercage = "Animal.husbandry..number.of.animals.per.cage",
```

```
        environmental = "Animal.husbandry..is.any.environmental.enrichment.reported.") %>%
    unique() %>%
    select(-StudyId) %>%
    mutate(lightcycle = case_when(lightcycle == "NR" ~ "NotReported",
        TRUE ~ "Reported"), animalspercage = case_when(animalspercage ==
        "NR" ~ "NotReported", TRUE ~ "Reported"), environmental = case_when(environmental ==
        "FALSE" ~ "NotReported", TRUE ~ "Reported")) %>%
    pivot_longer(cols = c(lightcycle, animalspercage, environmental),
        names_to = "variable") %>%
    group_by(variable, value) %>%
    summarise(count = n()) %>%
    pivot_wider(names_from = value, values_from = count) %>%
    mutate(Overall = Reported + NotReported, percentage = Reported/Overall *
        100, across(where(is.numeric), ~round(., 2))) %>%
    select(-c(NotReported, Reported, Overall))
```

Slice Protocol Measures

```
slice_data <- full_data_figures %>%
    select(StudyId, `Ca concentration in slicing solution` = "Ephys..calcium.concentration..mM..of.slici
        `Ca concentration in recording solution` = "Ephys..calcium.concentration..mM..of.recording.solut
        `Mg concentration in slicing solution` = "Ephys..magnesium.concentration..mM..of.slicing.solutio
        `Mg concentration in recording solution` = "Ephys..magnesium.concentration..mM..of.ACSF.used.fo
        `Plane slice taken from` = "Which.plane.was.the.slice.taken.from.",
        `Temperature slice left to recover at` = "Ephys..temperature...C..slices.left.to.recover.at",
        `Slice recovery time` = "How.long..minutes..were.brain.slices.left.to.recover.after.slicing.",
        `Kynurenic acid used` = "Ephys..was.kynurenic.acid.used.during.the.brain.dissection.",
        `Anaesthetised prior to sacrifice` = "Ephys..were.animals.anaesthetised.prior.to.decapitation."
        `Sacrifice time of day` = "Ephys..what.time.of.day.were.animals.sacrificed.") %>%
    unique() %>%
    select(-StudyId) %>%
    mutate(`Ca concentration in slicing solution` = case_when(`Ca concentration in slicing solution` ==
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Ca concentration in recording solution` = case_when(
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Mg concentration in slicing solution` = case_when(`Mg
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Mg concentration in recording solution` = case_when(
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Plane slice taken from` = case_when(`Plane slice take
        "Not reported" ~ "NotReported", TRUE ~ "Reported"), `Temperature slice left to recover at` = cas
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Slice recovery time` = case_when(`Slice recovery time
        "NR" ~ "NotReported", TRUE ~ "Reported"), `Kynurenic acid used` = case_when(`Kynurenic acid used
        "No" ~ "NotReported", TRUE ~ "Reported"), `Anaesthetised prior to sacrifice` = case_when(`Anaest
        "Not reported" ~ "NotReported", TRUE ~ "Reported"), `Sacrifice time of day` = case_when(`Sacrifi
        "Not reported" ~ "NotReported", TRUE ~ "Reported"), ) %>%
    pivot_longer(cols = c(`Ca concentration in slicing solution`,
        `Ca concentration in recording solution`, `Mg concentration in slicing solution`,
        `Mg concentration in recording solution`, `Plane slice taken from`,
        `Temperature slice left to recover at`, `Slice recovery time`,
        `Kynurenic acid used`, `Anaesthetised prior to sacrifice`,
        `Sacrifice time of day`), names_to = "variable") %>%
    group_by(variable, value) %>%
    summarise(count = n()) %>%
    pivot_wider(names_from = value, values_from = count) %>%
```

```
    mutate(Overall = Reported + NotReported, percentage = Reported/Overall *
        100, across(where(is.numeric), ~round(., 2))) %>%
    select(-c(NotReported, Reported, Overall))

slice_plot <- slice_data %>%
    ggplot(aes(x = reorder(variable, percentage), y = percentage)) +
    geom_bar(stat = "identity", fill = "#2683C6") + geom_text(aes(label = percentage),
    position = position_dodge(width = 0.9), hjust = 1.25, color = "white",
    size = 3.5) + theme(plot.title = element_text(hjust = 0.5)) +
    ggtitle("Studies(%) Reporing Different Slice Protocol Measures") +
    labs(x = "Protocol Measure", y = "Included (%)") + coord_flip() +
    scale_y_continuous(labels = abs, limits = c(0, 100))

slice_plot
```

Age and Sex Data

```
age_sex_data <- full_data_figures %>%
    filter(Outcome.category == "Electrophysiology") %>%
    select(StudyId, Outcome = "OutcomeLabel", Age = "What.age..weeks..were.animals.when.EPhys.outcomes.
        Sex = "Sex.of.animals.in.cohort") %>%
    unique()


sex_data <- age_sex_data %>%
    count(Sex) %>%
    mutate(overall = sum(n), percentage = n/overall * 100, across(where(is.numeric),
        ~round(., 2))) %>%
    ggplot(aes(x = Sex, y = percentage)) + ylim(0, 100) + geom_bar(stat = "identity",
    fill = "#2683C6") + geom_text(aes(label = percentage), vjust = 1.5,
    color = "white") + theme(plot.title = element_text(hjust = 0.5)) +
    ggtitle("Sex of Transgenic Animals Used") + labs(x = "Sex",
    y = "(%) Experiments Using Sex")

sex_data
```

```
age_count <- age_sex_data %>%
    count(Age)

age_data <- age_sex_data %>%
    mutate(Age = as.numeric(Age)) %>%
    ggplot(aes(x = Age)) + geom_histogram(aes(y = ..density..),
    fill = "#174E77") + theme(plot.title = element_text(hjust = 0.5)) +
    ggtitle("Age of Transgenic Animals Used") + labs(x = "Age (Weeks)",
    y = "Density of Experiments Using Age")

age_data
```

Publication Location

```
countries_included <- full_data_figures %>%
    select(region = "What.is.the.country.of.origin.of.the.corresponding.author",
        StudyId) %>%
    unique() %>%
    mutate(`Number of Studies` = 1) %>%
    select(-StudyId) %>%
    group_by(region) %>%
    summarise(`Number of Studies` = sum(`Number of Studies`)) %>%
    mutate(region = if_else(region == "United Kingdom", "UK",
        if_else(region == "United States", "USA", if_else(region ==
            "Korea, South", "South Korea", region))))


mapdata <- map_data("world") %>%
    left_join(countries_included, by = "region")

map1 <- mapdata %>%
    ggplot(aes(x = long, y = lat, group = group)) + geom_polygon(aes(fill = `Number of Studies`))
map1
```

Year Published

```
year_data <- full_data_figures %>%
    select(StudyId, Year) %>%
    unique() %>%
    mutate(`Number of Studies` = 1) %>%
    select(-StudyId) %>%
    group_by(Year) %>%
    summarise(`Number of Studies` = sum(`Number of Studies`)) %>%
    add_row(Year = 2023, `Number of Studies` = 0)

year_plot <- year_data %>%
    ggplot(aes(x = Year, y = `Number of Studies`, group = 1)) +
    geom_line() + geom_point() + geom_smooth(method = "lm", se = FALSE,
    col = "#174E77")

year_plot
```

Appendix 6: Code for Meta Analysis

Meta-Analysis

```
library(tidyverse)
library(kableExtra)
library(rmarkdown)
library(knitr)
library(janitor)
library(meta)
library(metafor)
library(readxl)
library(writexl)
```

Full data wrangling

```
full_data <- read.csv("full_data.csv") %>%
    filter(AnnotatorIdStr == "00374321-95a7-4a2a-a12c-af4c0f9e599b",
           NumberOfAnimals!= 0, #filters out studies with no n number
           StudyId != "6d537dfe-3d19-4dbf-b9af-bf9fad52bb4b")
```

#LTP data ## data wrangling to get the right data

```
LTPdata <- full_data %>%
    filter(Outcome.measure.category == "LTP")  #%>%
# write_xlsx('newLTP_data.xlsx') # use this data to add n
# numbers

LTP_data <- read_xlsx("LTP_data_newn.xlsx") %>%
    ## removes studies not needed
group_by(ExperimentID) %>%
    filter(any(ModelType == "model control") & any(ModelType ==
        "model"))
```

```
# selecting the right columns
LTP_model_data <- LTP_data %>%
    select(MatchId, StudyId, Author, OutcomeLabel, CohortId,
        ExperimentID, GreaterIsWorse, OutcomeResult, Outcome.measure.average.type,
        OutcomeError, Outcome.measure.error.type, ModelType,
        New_n, TimeInMinute, Sex.of.animals.in.cohort, What.age..weeks..were.animals.when.EPhys.outcomes
```

LTP 30 minutes

```
## LTP at 30 min
LTP_model_data_30 <- LTP_model_data %>%
  filter(TimeInMinute >29 & TimeInMinute< 31) %>% # first look at LTP at 30 min
  mutate(TimeInMinute = 30) %>%
  mutate(SD = ifelse(Outcome.measure.error.type == "SD",
                     OutcomeError, OutcomeError * sqrt(New_n)))%>%
  mutate(SEM = ifelse(Outcome.measure.error.type == "SEM",
                      OutcomeError, SD / sqrt(New_n))) %>%
  unique() %>% select(-OutcomeError) %>% # remove OutcomeError column
  mutate(Direction= ifelse(GreaterIsWorse ==TRUE, 1, -1)) %>%
  mutate(New_n = as.numeric(New_n)) %>%
  mutate(OutcomeResult = as.numeric(OutcomeResult)) %>%
  mutate(SEM = as.numeric(SEM)) %>%
  mutate(SD = as.numeric(SD))
```

Widen data set (30)

```
LTP_model_data_30_wide <- LTP_model_data_30 %>%
    pivot_wider(names_from = ModelType, values_from = c(New_n,
        SD, SEM, OutcomeResult, CohortId), values_fn = list) %>%
    unnest(cols = everything())
```

Calculating effect sizes (30)

```
LTP_model_data_30_final <- LTP_model_data_30_wide %>%
  mutate(combined_n = `New_n_model control` + New_n_model) %>%
  mutate(SPooled = sqrt(((`New_n_model control`- 1) * `SD_model control`^2 +
                           (New_n_model - 1) * SD_model^2)/ (combined_n-2))) %>%
  mutate(SMD_step1 = (`OutcomeResult_model control`-`OutcomeResult_model`)
           /SPooled) %>%
  mutate(SMD_step2 = (1 - (3/(4*combined_n-9)))) %>%
  mutate(SMD_ES = SMD_step1 * SMD_step2 * Direction) %>% #equation 15
  mutate(SMD_ES_SE = sqrt(
    combined_n/(New_n_model  * `New_n_model control`) +
           SMD_ES^2/(2*(combined_n - 3.94)))) %>% #equation 16
  mutate(weight = 1/(SMD_ES_SE^2)) %>% #equation 20
  mutate(weighted_ES = SMD_ES * weight) %>% #equation 21
  group_by(StudyId) %>%
  mutate(id = cur_group_id()) %>%
  select(id, everything())
```

Random effects meta-analysis (30)

```
ma_results_ltp30 <- metagen(
  `SMD_ES`, # specify the effect size for each experiment
  `SMD_ES_SE`, # specify the variable that contains the standard error for each experiment
  sm = "SMD",
```

```
  data = LTP_model_data_30_final, # specify the data set
  studlab = id, # specify the study labels
  comb.random = TRUE, # specify a random effects model
  comb.fixed = FALSE,
  method.tau = "REML") # specify which method is used to estimate the between-study variance

summary(ma_results_ltp30)
```

Forest Plot (30)

```
forest_plot_30 <- forest(
  ma_results_ltp30, # specify the meta-analysis to plot
  sortvar = TE, # sort the data according to effect size
  comb.fixed = FALSE, # do not plot the fixed effect estimate
  comb.random = TRUE, # plot the random effects estimate
  xlab = "SMD", # specify the x axis label
  smlab = "Standardised Mean Difference", # specify the effect size label
  label.right = "Increased LTP", # specify the graph label on right side of plot
  label.left = "Decreased LTP", # specify the graph label on left side of plot
  fontsize = 18, # specify the size of text (in points)
  plotwidth = "15cm", # specify the width of the plotting region
  digits = 1, # specify minimal number of significant digits for treatment effects
  digits.se = 1, # specify minimal number of significant digits for standard errors
  col.square="lightblue",
  leftcols = c("id"),
  leftlabs = c("id"),
  col.diamond = "darkblue",
  spacing=1.2
)
```

Publication Bias (30)

```
# Define fill colors for contour
col.contour = c("gray75", "gray85", "gray95")

# Generate funnel plot (we do not include study labels here)
funnel.meta(ma_results_ltp30, xlim = c(-8, 2), #took the SMD from the forest plot
            contour = c(0.9, 0.95, 0.99),
            col.contour = col.contour)

# Add a legend
legend(x = -7.5, y = 0.01,
       legend = c("p < 0.1", "p < 0.05", "p < 0.01"),
       fill = col.contour)

# Add a title
title("Funnel Plot of LTP after 30 min")
```

```
## Adds columns to the ma_results_ltp30 to use in the Putejovsky analysis
ma_results_ltp30$n.e = LTP_model_data_30_final$New_n_model
ma_results_ltp30$n.c = LTP_model_data_30_final$`New_n_model control`

# Pustejovsky
metabias(ma_results_ltp30, method.bias = "Pustejovsky")
```

Trim and Fill (30)

Needed to load to find outliers https://raw.githubusercontent.com/MathiasHarrer/dmetar/master/R/find.outliers.R

```
## Trim Fill
tf_ltp30 <- trimfill(ma_results_ltp30)
summary(tf_ltp30)


# Define fill colors for contour
contour <- c(0.9, 0.95, 0.99)
col.contour <- c("gray75", "gray85", "gray95")
ld <- c("p < 0.1", "p < 0.05", "p < 0.01")

# Use 'par' to create two plots in one row (row, columns)
par(mfrow=c(1,2))

# Contour-enhanced funnel plot (full data)
funnel.meta(tf_ltp30,
            xlim = c(-8, 8), contour = contour,
            col.contour = col.contour)
legend(x = -7.5, y = 0,
       legend = ld, fill = col.contour)
title("Funnel Plot (Trim & Fill Method)")


# find outliers
find.outliers(ma_results_ltp30) ## have to load a package (chap 5.2 in handbook)

# filter out outliers and rerun meta and funnel plot
ltp30_nooutliers <- LTP_model_data_30_final %>%
  filter(id != 10,
         id != 5,
         id != 3)

ma_results_ltp30_nooutliers <- metagen(
  `SMD_ES`, # specify the effect size for each experiment
  `SMD_ES_SE`, # specify the variable that contains the standard error for each experimen
  sm = "SMD",
  data = ltp30_nooutliers, # specify the data set
  studlab = id, # specify the study labels
  comb.random = TRUE, # specify a random effects model
  comb.fixed = FALSE,
  method.tau = "REML") # specify which method is used to estimate the between-study variance
```

```r
summary(ma_results_ltp30_nooutliers) ## reduces heterogenity to 25%

tf_ltp30_nooutliers <- trimfill(ma_results_ltp30_nooutliers)
summary(tf_ltp30_nooutliers)


# Define fill colors for contour
contour <- c(0.9, 0.95, 0.99)
col.contour <- c("gray75", "gray85", "gray95")
ld <- c("p < 0.1", "p < 0.05", "p < 0.01")


# Contour-enhanced funnel plot (outliers removed)
funnel.meta(tf_ltp30_nooutliers,
            xlim = c(-8, 8), contour = contour,
            col.contour = col.contour)
legend(x = -7.5, y = 0.01,
       legend = ld, fill = col.contour)
title("Funnel Plot (Trim & Fill Method)
      -Outliers Removed")
```

LTP 60 minutes

```r
## LTP at 30 min
LTP_model_data_60 <- LTP_model_data %>%
  filter(TimeInMinute >59 & TimeInMinute< 61) %>% # first look at LTP at 30 min
  mutate(TimeInMinute = 60) %>%
  mutate(SD = ifelse(Outcome.measure.error.type == "SD",
                     OutcomeError, OutcomeError * sqrt(New_n)))%>%
  mutate(SEM = ifelse(Outcome.measure.error.type == "SEM",
                      OutcomeError, SD / sqrt(New_n))) %>%
  unique() %>% select(-OutcomeError) %>% # remove OutcomeError column
  mutate(Direction= ifelse(GreaterIsWorse ==TRUE, 1, -1)) %>%
  mutate(New_n = as.numeric(New_n)) %>%
  mutate(OutcomeResult = as.numeric(OutcomeResult)) %>%
  mutate(SEM = as.numeric(SEM)) %>%
  mutate(SD = as.numeric(SD))
```

Widen data set(60)

```r
LTP_model_data_60_wide <- LTP_model_data_60 %>%
    pivot_wider(names_from = ModelType, values_from = c(New_n,
        SD, SEM, OutcomeResult, CohortId), values_fn = list) %>%
    unnest(cols = everything())
```

Calculating effect sizes (60)

```
LTP_model_data_60_final <- LTP_model_data_60_wide %>%
  mutate(combined_n = `New_n_model control` + New_n_model) %>%
  mutate(SPooled = sqrt(((`New_n_model control`- 1) * `SD_model control`^2 +
                           (New_n_model - 1) * SD_model^2)/ (combined_n-2))) %>%
  mutate(SMD_step1 = (`OutcomeResult_model control`-`OutcomeResult_model`)
           /SPooled) %>%
  mutate(SMD_step2 = (1 - (3/(4*combined_n-9)))) %>%
  mutate(SMD_ES = SMD_step1 * SMD_step2 * Direction) %>% #equation 15
  mutate(SMD_ES_SE = sqrt(
    combined_n/(New_n_model  * `New_n_model control`) +
           SMD_ES^2/(2*(combined_n - 3.94)))) %>% #equation 16
  mutate(weight = 1/(SMD_ES_SE^2)) %>% #equation 20
  mutate(weighted_ES = SMD_ES * weight) %>% #equation 21
  group_by(StudyId) %>%
  mutate(id = cur_group_id()) %>%
  select(id, everything())
```

Random effects meta-analysis (60)

```
ma_results_ltp60 <- metagen(
  `SMD_ES`, # specify the effect size for each experiment
  `SMD_ES_SE`, # specify the variable that contains the standard error for each experimen
  sm = "SMD",
  data = LTP_model_data_60_final, # specify the data set
  studlab = id, # specify the study labels
  comb.random = TRUE, # specify a random effects model
  comb.fixed = FALSE,
  method.tau = "REML") # specify which method is used to estimate the between-study variance

summary(ma_results_ltp60)
```

Forest Plot (60)

```
forest_plot_60 <- forest(
  ma_results_ltp60, # specify the meta-analysis to plot
  sortvar = TE, # sort the data according to effect size
  comb.fixed = FALSE, # do not plot the fixed effect estimate
  comb.random = TRUE, # plot the random effects estimate
  xlab = "SMD", # specify the x axis label
  smlab = "Standardised Mean Difference", # specify the effect size label
  label.right = "Increased LTP", # specify the graph label on right side of plot
  label.left = "Decreased LTP", # specify the graph label on left side of plot
  fontsize = 18, # specify the size of text (in points)
  plotwidth = "15cm", # specify the width of the plotting region
  digits = 1, # specify minimal number of significant digits for treatment effects
  digits.se = 1, # specify minimal number of significant digits for standard errors
```

```
  col.square="lightblue",
  leftcols = c("id"),
  leftlabs = c("id"),
  col.diamond = "darkblue",
  spacing=1.2
)
```

#I/O data

```
IO_data <- full_data %>%
    filter(OutcomeLabel == "I/O")  #%>%
# write_xlsx('IO_data.xlsx')

IO_data <- read_xlsx("IO_data_newn.xlsx") %>%
    group_by(ExperimentID) %>%
    filter(any(ModelType == "model control") & any(ModelType ==
        "model"))
```

```
# selecting the right columns
IO_model_data <- IO_data %>%
    select(MatchId, StudyId, Author, OutcomeLabel, CohortId,
        ExperimentID, GreaterIsWorse, OutcomeResult, Outcome.measure.average.type,
        OutcomeError, Outcome.measure.error.type, ModelType,
        New_n, TimeInMinute, Sex.of.animals.in.cohort, What.age..weeks..were.animals.when.EPhys.outcomes
    mutate(TimeInMinute = "1")
```

calculating SD and SEM

```
IO_model_data <- IO_model_data %>%
  mutate(SD = ifelse(Outcome.measure.error.type == "SD", OutcomeError,
                     OutcomeError * sqrt(New_n)))%>%
  mutate(SEM = ifelse(Outcome.measure.error.type == "SEM",
                      OutcomeError, SD / sqrt(New_n))) %>%
  unique() %>% select(-OutcomeError) %>% # remove OutcomeError column
  mutate(Direction= ifelse(GreaterIsWorse ==TRUE, 1, -1)) %>%
  mutate(New_n = as.numeric(New_n)) %>%
  mutate(OutcomeResult = as.numeric(OutcomeResult)) %>%
  mutate(SEM = as.numeric(SEM)) %>%
  mutate(SD = as.numeric(SD))
```

Widen data set I/O

```
IO_model_data_wide <- IO_model_data %>%
    pivot_wider(names_from = ModelType, values_from = c(New_n,
        SD, SEM, OutcomeResult, CohortId), values_fn = list) %>%
    unnest(cols = everything())
```

Calculating effect sizes I/O

```
IO_model_data_final <- IO_model_data_wide %>%
  mutate(combined_n = `New_n_model control` + New_n_model) %>%
  mutate(SPooled = sqrt(((`New_n_model control`- 1) * `SD_model control`^2 +
                           (New_n_model - 1) * SD_model^2)/ (combined_n-2))) %>%
  mutate(SMD_step1 = (`OutcomeResult_model control`-`OutcomeResult_model`)
           /SPooled) %>%
  mutate(SMD_step2 = (1 - (3/(4*combined_n-9)))) %>%
  mutate(SMD_ES = SMD_step1 * SMD_step2 * Direction) %>% #equation 15
  mutate(SMD_ES_SE = sqrt(
    combined_n/(New_n_model  * `New_n_model control`) +
           SMD_ES^2/(2*(combined_n - 3.94)))) %>% #equation 16
  mutate(weight = 1/(SMD_ES_SE^2)) %>% #equation 20
  mutate(weighted_ES = SMD_ES * weight) %>% #equation 21
  group_by(StudyId) %>%
  mutate(id = cur_group_id()) %>%
  select(id, everything())
```

Random effects meta-analysis I/O

```
ma_results_IO <- metagen(
  `SMD_ES`, # specify the effect size for each experiment
  `SMD_ES_SE`, # specify the variable that contains the standard error for each experiment
  sm = "SMD",
  data = IO_model_data_final, # specify the data set
  studlab = id, # specify the study labels
  comb.random = TRUE, # specify a random effects model
  comb.fixed = FALSE,
  method.tau = "REML") # specify which method is used to estimate the between-study variance

summary(ma_results_IO)


smd_IO <- ma_results_IO %>%
  as.data.frame() %>%
  select("SMD" = "TE", "id" = "studlab") ## gets the SMD into a data frame
```

Forest Plot I/O

```
forest_plot_IO <- forest(
  ma_results_IO, # specify the meta-analysis to plot
  sortvar = TE, # sort the data according to effect size
  comb.fixed = FALSE, # do not plot the fixed effect estimate
  comb.random = TRUE, # plot the random effects estimate
  xlab = "SMD", # specify the x axis label
  smlab = "Standardised Mean Difference", # specify the effect size label
  label.right = "Increased I/O", # specify the graph label on right side of plot
```

```
  label.left = "Decreased I/O", # specify the graph label on left side of plot
  fontsize = 18, # specify the size of text (in points)
  plotwidth = "15cm", # specify the width of the plotting region
  digits = 1, # specify minimal number of significant digits for treatment effects
  digits.se = 1, # specify minimal number of significant digits for standard errors
  col.square="lightblue",
  leftcols = c("id"),
  leftlabs = c("id"),
  col.diamond = "darkblue",
  spacing=1.2
)
```

Other Data needed

```
PPF_data <- full_data %>%
  filter(OutcomeLabel == c("PPF", "PPR")) %>%
  group_by(ExperimentID) %>%
  filter(any(ModelType == "model control") &
           any(ModelType == "model")) %>%
  select(StudyId, Author, OutcomeLabel,
         CohortId, ExperimentID, GreaterIsWorse, OutcomeResult,
         Outcome.measure.average.type, OutcomeError,
         Outcome.measure.error.type, ModelType, Number.of.animals.in.cohort,
         TimeInMinute, Sex.of.animals.in.cohort,
         What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed.) %>%
  filter(TimeInMinute == "50")


LTP_90_data <- full_data %>%
  filter(OutcomeLabel == "LTP") %>%
  group_by(ExperimentID) %>%
  filter(any(ModelType == "model control") &
           any(ModelType == "model")) %>%
  select(StudyId, Author, OutcomeLabel,
         CohortId, ExperimentID, GreaterIsWorse, OutcomeResult,
         Outcome.measure.average.type, OutcomeError,
         Outcome.measure.error.type, ModelType, Number.of.animals.in.cohort,
         TimeInMinute, Sex.of.animals.in.cohort,
         What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed.) %>%
  filter(TimeInMinute >89 & TimeInMinute< 91) %>% # first look at LTP at 30 min
  mutate(TimeInMinute = 90)


LTD_30_data <- full_data %>%
  filter(OutcomeLabel == "LTD") %>%
  group_by(ExperimentID) %>%
  filter(any(ModelType == "model control") &
           any(ModelType == "model")) %>%
  select(StudyId, Author, OutcomeLabel,
         CohortId, ExperimentID, GreaterIsWorse, OutcomeResult,
         Outcome.measure.average.type, OutcomeError,
```

```
         Outcome.measure.error.type, ModelType, Number.of.animals.in.cohort,
         TimeInMinute, Sex.of.animals.in.cohort,
         What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed.) %>%
  filter(TimeInMinute >29 & TimeInMinute< 31) %>% # first look at LTP at 30 min
  mutate(TimeInMinute = 30)

LTD_60_data <- full_data %>%
  filter(OutcomeLabel == "LTD") %>%
  group_by(ExperimentID) %>%
  filter(any(ModelType == "model control") &
           any(ModelType == "model")) %>%
  select(StudyId, Author, OutcomeLabel,
         CohortId, ExperimentID, GreaterIsWorse, OutcomeResult,
         Outcome.measure.average.type, OutcomeError,
         Outcome.measure.error.type, ModelType, Number.of.animals.in.cohort,
         TimeInMinute, Sex.of.animals.in.cohort,
         What.age..weeks..were.animals.when.EPhys.outcomes.were.assessed.) %>%
  filter(TimeInMinute >59 & TimeInMinute< 61) %>% # first look at LTP at 30 min
  mutate(TimeInMinute = 60)
```

References

Brownlee, C., Shannon, G. & Bielekova, B. (2021), 'Ending the reproducibility crisis', *Issues in Science and Technology* **38**(1), 82–88.
URL: *https://www.jstor.org/stable/27092121*

Chapman, P. F., White, G. L., Jones, M. W., Cooper-Blacketer, D., Marshall, V. J., Irizarry, M., Younkin, L., Good, M. A., Bliss, T. V. P., Hyman, B. T., Younkin, S. G. & Hsiao, K. K. (1999), 'Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice', *Nature Neuroscience* **2**, 271–276.

Choi, W., Hong, H., Adams, R. & Bonett, D. (2018), 'Testing small study effects in multivariate meta-analysis', *Biometrics* **74**(4), 1419–1429.

Citri, A. & Malenka, R. C. (2007), 'Synaptic plasticity: Multiple forms, functions, and mechanisms', *Neuropsychopharmacology* **33**, 18–41.
URL: *https://www.nature.com/articles/1301559*

D'Amelio, M., Cavallucci, V., Middei, S., Marchetti, C., Pacioni, S., Ferri, A., Diamantini, A., De Zio, D., Carrara, P., Battistini, L., Moreno, S., Bacci, A., Ammassari-Teule, M., Marie, H. & Cecconi, F. (2010), 'Caspase-3 triggers early synaptic dysfunction in a mouse model of alzheimer's disease', *Nature Neuroscience* **14**, 69–76.

Dunn, A. G., Arachi, D., Hudgins, J. & Tsafnat, G. (2020), 'Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial', *Journal of Clinical Epidemiology* **121**, 68–75.

Egan, K., Vesterinen, H., Beglopoulos, V., Sena, E. & Macleod, M. (2016), 'From a mouse: systematic analysis reveals limitations of experiments testing interventions in alzheimer's disease mouse models', *Evidence-based Preclinical Medicine* **3**, e00015.

Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997), 'Bias in meta-analysis detected by a simple, graphical test', *BMJ (Clinical research ed.)* **315**(7109), 629–634.

Elder, G. A., Gama Sosa, M. A. & De Gasperi, R. (2010), 'Transgenic mouse models of alzheimer's disease', *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* **77**, 69–81.
URL: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2925685/*

Fitzjohn, S. M., Morton, R. A., Kuenzi, F., Rosahl, T. W., Shearman, M., Lewis, H., Smith, D., Reynolds, D. S., Davies, C. H., Collingridge, G. L. & Seabrook, G. R. (2001), 'Age-related impairment of synaptic transmission but normal long-term potentiation in transgenic mice that overexpress the human app695swe mutant form of amyloid precursor protein', *The Journal of Neuroscience* **21**, 4691–4698.

Golde, T. E. (2022), 'Disease-modifying therapies for alzheimer's disease: More questions than answers', *Neurotherapeutics* .

Hair, K. (2021), Developing automated metaresearch approaches in the preclinical Alzheimer's disease literature, Phd thesis, Edinburgh.

Hair, K., Wilson, E., Wong, C., Tsang, A., Macleod, M. & Bannach-Brown, A. (2022), 'Systematic online living evidence summaries: emerging tools to accelerate evidence synthesis', *OSF Preprints* .

Hardy, J. & Higgins, G. (1992), 'Alzheimer's disease: the amyloid cascade hypothesis', *Science* **256**, 184–185.

Harrer, M., Cuijpers, P., A, F. T. & Ebert, D. D. (2021), *Doing Meta-Analysis With R: A Hands-On Guide*, 1st edn, Chapman Hall/CRC Press, Boca Raton, FL and London.

Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., Yang, F. & Cole, G. (1996), 'Correlative memory deficits, abeta elevation, and amyloid plaques in transgenic mice', *Science* **274**(5284), 99–102.

Koffie, R. M., Hyman, B. T. & Spires-Jones, T. L. (2011), 'Alzheimer's disease: synapses gone cold', *Molecular neurodegeneration* **6**(1), 63.

Lazic, S. E., Clarke-Williams, C. & Munafo, M. (2018), 'What exactly is 'n' in cell culture and animal experiments?', *PLoS biology* **16**(6), e2005282.

Marchetti, C. & Marie, H. (2011), 'Hippocampal synaptic plasticity in alzheimer's disease: what have we learned so far from transgenic models?', *Reviews in the Neurosciences* **22**.

Mehla, J., Lacoursiere, S. G., Lapointe, V., McNaughton, B. L., Sutherland, R. J., McDonald, R. J. & Mohajerani, M. H. (2019), 'Age-dependent behavioral and biochemical characterization of single app knock-in mouse (appnl-g-f/nl-g-f) model of alzheimer's disease', *Neurobiology of Aging* **75**, 25–37.
URL: *https://www.sciencedirect.com/science/article/pii/S0197458018303907*

Mucke, L., Masliah, E., Yu, G. Q., Mallory, M., Rockenstein, E. M., Tatsuno, G., Hu, K., Kholodenko, D., Johnson-Wood, K. & McConlogue, L. (2000), 'High-level neuronal expression of aβ1–42 in wild-type human amyloid protein precursor transgenic mice: Synaptotoxicity without plaque formation', *The Journal of Neuroscience* **20**(11), 4050–4058.

Organisation, W. H. (2021), 'Dementia'.
URL: *https://www.who.int/news-room/fact-sheets/detail/dementia*

Othman, M. Z., Hassan, Z. & Che Has, A. T. (2022), 'Morris water maze: a versatile and pertinent tool for assessing spatial learning and memory', *Experimental Animals* .

Pustejovsky, J. E. & Rodgers, M. A. (2019), 'Testing for funnel plot asymmetry of standardized mean differences', *Research Synthesis Methods* **10**(1), 57–71.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1332*

Scheltens, P. (2021), 'Alzheimer's disease', *The Lancet* **397**, 1577–1590.
URL: *https://www.sciencedirect.com/science/article/abs/pii/S0140673620322054*

Simonsohn, U., Simmons, J. P. & Nelson, L. D. (2015), 'Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to ulrich and miller (2015)', *Journal of Experimental Psychology: General* **144**(6), 1146–1152.

Sri, S., Pegasiou, C.-M., Cave, C. A., Hough, K., Wood, N., Gomez-Nicola, D., Deinhardt, K., Bannerman, D. M., Perry, V. H. & Vargas-Caballero, M. (2019), 'Emergence of synaptic and cognitive impairment in a mature-onset app mouse model of alzheimer's disease', *Acta neuropathologica communications* **7**(1).

Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., Antonic, A., Howells, D. W. & Macleod, M. R. (2014), 'Systematic review and stratified meta-analysis of the efficacy of interleukin-1 receptor antagonist in animal models of stroke', *Journal of Neuroimmunology* **269**(1-2), 16–22.

Villegas, S., Roda, A., Serra-Mir, G., Montoliu-Gaya, L. & Tiessler, L. (2022), 'Amyloid-beta peptide and tau protein crosstalk in alzheimer's disease', *Neural Regeneration Research* **17**, 1666.

Xiong, G., Metheny, H., Johnson, B. N. & Cohen, A. S. (2017), 'A comparison of different slicing planes in preservation of major hippocampal pathway fibers in the mouse', *Frontiers in Neuroanatomy* **11**, 107.

Yoshiyama, Y., Higuchi, M., Zhang, B., Huang, S.-W., Iwata, N. & Saido, T. C. (2015), 'Single APP knock-in mouse models of Alzheimer's disease', *Nature Neuroscience* **18**(4), 617–626.