THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Developing automated meta-research approaches in the preclinical Alzheimer's disease literature

Kaitlyn Hair

PhD Thesis

The University of Edinburgh

2021

DECLARATION

I declare that this thesis has been composed by me and the work presented here is that of my own, unless clearly stated within each section. I confirm that this work has not been submitted for any other degree or professional qualification.

Kaitlyn Hair Date: 02/09/2021

ACKNOWLEDGEMENTS

I would first like to acknowledge all my collaborators, and funders who made the work described in this thesis possible.

I would like to thank the wonderful team of reviewers who contributed to the screening and categorisation of research (Chapter 2) and/or the crowdsourced systematic review project (Chapter 6): Isabel Lefevre, Farah Francis, Lukas Bohsung, Bettina Braun, Isabel Seiffert, Nadia Soliman, Aishwarya Chinchanikar, Vincent Castagné, Kim Wever, Coco Zhang, Berhnard Voelkl, Anna Zemel, Darlyn Ranis, Sara Steele, Zsanett Bahor, Jennifer Mollon, Daniel McLoone, Louis Dwomoh, Rita Bertani, Ross Davidson, Alex Davidson, Hayley Leake, Sabrina Gauchan, Ann Marie Waldron, Heidrun Potschka, Fraser Sneden, Robert Humphries, Muhammad Inamekhuda, Annesha Sil, Juan Diego Pita Almenar, Malcolm Macleod, Gillian Currie, Anna Davidson, and Emily Sena.

I would also like to thank Kim Wever, Darlyn Ranis, Emily Wheater, and Alexandra Bannach-Brown for providing the gold standard deduplicated systematic search datasets used to develop the automated deduplication tool in Chapter 4.

I am grateful to Dr Juan Pita Almenar and the rest of the Janssen electrophysiology team for all your expert guidance, and for your patience in helping me get to grips with electrophysiology experiments in the lab.

I am thankful to have received a junior Alzheimer's Research UK Scotland Network grant to hire my summer student Megan McManus. Thank you, Meg, for all your hard work extracting data from publications for the synaptic plasticity and transmission systematic review (Chapter 5).

I would also like to extend a big thank you to the Innovative Medicines Initiative for funding my PhD research.

I would also like to thank my colleagues, friends, and family who have supported me along the way. Firstly, I would like to thank my supervisor Dr Emily Sena. You are not only an incredible mentor, but an inspirational person and a valued friend. I am incredibly grateful for the trust you have placed in me over the years and for your continued support, guidance, and enthusiasm. Thank you to Professor Malcolm Macleod for the many words of wisdom you have given me along the way. I am incredibly grateful to you for taking a chance on me and giving me the opportunity to join the amazing CAMARADES team in the first place. I can't imagine how different life might have been if I had taken a different path – I might have never left Glasgow… or I might have never made it into a British Airways lounge! For this, and for every science and/or gin-related experience in Edinburgh, London, Belgium, Berlin, and beyond, thank you. To everyone in the CAMARADES team both past and present - thank you all for making this journey full of laughs and learning, even throughout a global pandemic! Special thanks to Emma Wilson and Dr Charis Wong for your help in proofreading thesis chapters.

To Ezgi, my PhD partner in crime, thank you for being a fantastic friend, conference buddy, and travel companion. There are too many hilarious, cringe-worthy, and stressful moments to mention, but I'm so glad we went through all of them together.

Thank you to my Mum for her unwavering support and belief in me, no matter what I do.

Last, but certainly not least, Chris. Thank you for supporting me through tough times, bringing me endless tea, and for always listening (even if you have no idea what I'm talking about). You have been my rock throughout this journey, and I am so grateful to you for everything.

CONFERENCE PARTICIPATION

Lightning Talk: A "Living" Evidence Framework for Alzheimer's Disease Studies (Invited oral presentation) at *Meta x Data Meeting, BIH Quest Centre, Berlin, Germany,* October 2019 (Invited oral presentation)
Available at: https://bit.ly/3xEQ6HU

A living systematic review of animal models of Alzheimer's disease*, Alzheimer's Research UK Scotland Network Centre Meeting, St Andrews, UK*, August 2019 (Oral presentation)

Selfish reasons to work reproducibly, *Edinburgh ReproducibiliTea Journal Club, Online,* November 2020 (Invited oral presentation)

A "Living" Evidence Framework for Alzheimer's Disease studies at the *EQUATOR Conference, Berlin, Germany, February 2020* (Oral presentation, best scoring abstract and travel grant awarded)

Team Credibility Prize Presentation (Invited oral presentation as part of the BNA Team Credibility Award) at *British Neuroscience Association Conference, Online,* April 2021. (Invited oral presentation)

Identifying duplicate publications with the ASySD (Automated Systematic Search De-duplication) tool, *Evidence Synthesis & Meta-Analysis in R Conference, Online,* January 2021, (Oral presentation).
Available at: https://bit.ly/3jWs1Yv

Translational failure in preclinical research: how open science can help, *GRECO project: Inspiring researchers on the use of Open Science in their projects, Online, May 2021* (Invited oral presentation)
Available at: https://bit.ly/3yCJ1c7

Building a Systematic, Online, Living Evidence Summary of COVID-19 Research, *International Collaboration for the Automation of Systematic Reviews, Online,* May 2021 (Invited oral presentation)

PUBLICATIONS AND PREPRINTS

(1) Wang, Q., Liao, J., **Hair, K**., Bannach-Brown, A., Bahor, Z., Currie, G. L., ... & Macleod, M. R. (2018). Estimating the statistical performance of different approaches to meta-analysis of data from animal studies in identifying the impact of aspects of study design. Biorxiv, 256776.

(2) Currie, G. L., Angel-Scott, H. N., Colvin, L., Cramond, F., **Hair, K**., Khandoker, L., ... & Sherratt, N. (2019). Animal models of chemotherapy-induced peripheral neuropathy: A machine-assisted systematic review and meta-analysis. PLoS biology, 17(5), e3000243.

(3) Macleod, M. R., Tanriver-Ayder, E., **Hair, K**., & Sena, E. (2019). Design of Meta-Analysis Studies. *Handb Exp Pharmacol*.

(4) **Hair, K**., Macleod, M. R., & Sena, E. S. and on behalf of the IICARus Collaboration (2019). A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Research integrity and peer review*, *4*(1), 12.

(5) **NPQIP Collaborative group**, & Study steering committee. (2019). Did a change in Nature journals' editorial policy for life sciences research improve reporting?. *BMJ Open Science*, 3(1), e000035.

(6) Macleod, M., **Hair, K**., Ezgi, T. A., Craske, D., Shinton, S., Campbell, A., … Thomson, L. (2021). Results of a randomised controlled trial comparing two different incentives to improve survey response rates. https://doi.org/10.31235/osf.io/sm36a

(7) Wang, Q., **Hair, K**., Macleod, M., Currie, G., Bahor, Z., Sena, E., & Liao, J. (2021). Protocol for an analysis of *in vivo* reporting standards by journal, institution and funder. https://doi.org/10.31222/osf.io/cjxtf

(8) **Hair, K.**, Bahor, Z., Macleod, M., Liao, J., & Sena, E. S. (2021). The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews.

(9) Bannach-Brown A, **Hair K**, Bahor Z*, et al., (2021)* Technological advances in preclinical meta-research. *BMJ Open Science*, DOI: 10.1136/bmjos-2020-100131

(10) Whaley, P., Blaauboer, B. J., Brozek, J., Hubal, E. A. C., **Hair, K**., Kacew, S., ... & Woodruff, T. J. (2021). Improving the quality of toxicology and environmental health systematic reviews: What journal editors can do. *ALTEX-Alternatives to animal experimentation*, *38*(3), 513-522.

(11) **Hair K**, Sena E, Wilson E, Currie G, Macleod M, Bahor Z, Sena C, et al. (2021) Building a Systematic Online Living Evidence Summary of COVID-19 Research. *Journal of European Association for Health Information and Libraries*. Available from: http://ojs.eahil.eu/ojs/index.php/JEAHIL/article/view/465

LIST OF TABLES

LIST OF FIGURES

CONTENTS

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
APP	Amyloid precursor protein
CAMARADES	Collaborative Approach to Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies
I/O	Input-output relationship
LTD	Long-term depression
LTP	Long-term potentiation
MAPT	Microtubule associated protein tau
MIO	Model(s), Intervention(s), Outcome(s)
ML	Machine learning
MWM	Morris water maze
OFT	Open field test
PICO	Population, Intervention, Comparator, Outcome
PPF	Paired pulse facilitation
Regex(es)	Regular expression(s)
RoB	Risk of bias
SMD	Standardised mean difference
SRs	Systematic reviews
SyRF	The systematic review facility (our in-house systematic review platform) https://syrf.org.uk/

ABSTRACT

Alzheimer's disease is a devastating neurodegenerative disorder for which there is no cure. A crucial part of the drug development pipeline involves testing therapeutic interventions in animal disease models. However, promising findings in preclinical experiments have not translated into clinical trial success. Reproducibility has often been cited as a major issue affecting biomedical research, where experimental results in one laboratory cannot be replicated in another. By using meta-research (research on research) approaches such as systematic reviews, researchers aim to identify and summarise all available evidence relating to a specific research question. By conducting a meta-analysis, researchers can also combine the results from different experiments statistically to understand the overall effect of an intervention and to explore reasons for variations seen across different publications. Systematic reviews of the preclinical Alzheimer's disease literature could inform decision making, encourage research improvement, and identify gaps in the literature to guide future research. However, due to the vast amount of potentially useful evidence from animal models of Alzheimer's disease, it remains difficult to make sense of and utilise this data effectively. Systematic reviews are common practice within evidence based medicine, yet their application to preclinical research is often limited by the time and resources required. In this thesis, I develop, build-upon, and implement automated meta-research approaches to collect, curate, and evaluate the preclinical Alzheimer's literature. I searched several biomedical databases to obtain all research relevant to Alzheimer's disease. I developed a novel deduplication tool to automatically identify and remove duplicate publications identified across different databases with minimal human effort. I trained a crowd of reviewers to annotate a subset of the publications identified and used this data to train a machine learning algorithm to screen through the remaining publications for relevance. I developed text-mining tools to extract model, intervention, and treatment information from publications and I improved existing automated tools to extract reported measures to reduce the risk of bias. Using these tools, I created a categorised database of research in transgenic Alzheimer's disease animal models and created a visual summary of this dataset on an interactive, openly accessible online platform. Using the techniques described, I also identified relevant publications within the categorised dataset to perform systematic reviews of two key outcomes of interest in transgenic Alzheimer's disease

models: (1) synaptic plasticity and transmission in hippocampal slices and (2) motor activity in the open field test.

Over 400,000 publications were identified across biomedical research databases, with 230,203 unique publications. In a performance evaluation across different preclinical datasets, the automated deduplication tool I developed could identify over 97% of duplicate citations and a had an error rate similar to that of human performance. When evaluated on a test set of publications, the machine learning classifier trained to identify relevant research in transgenic models performed was highly sensitive (captured 96.5% of relevant publications) and excluded 87.8% of irrelevant publications. Tools to identify the model(s) and outcome measure(s) within the full-text of publications may reduce the burden on reviewers and were found to be more sensitive than searching only the title and abstract of citations. Automated tools to assess risk of bias reporting were highly sensitive and could have the potential to monitor research improvement over time. The final dataset of categorised Alzheimer's disease research contained 22,375 publications which were then visualised in the interactive web application. Within the application, users can see how many publications report measures to reduce the risk of bias and how many have been classified as using each transgenic model, testing each intervention, and measuring each outcome. Users can also filter to obtain curated lists of relevant research, allowing them to perform systematic reviews at an accelerated pace with reduced effort required to search across databases, and a reduced number of publications to screen for relevance. Both systematic reviews and meta-analyses highlighted failures to report key methodological information within publications. Poor transparency of reporting limited the statistical power I had to understand the sources of between-study variation. However, some variables were found to explain a significant proportion of the heterogeneity. Transgenic animal model had a significant impact on results in both reviews. For certain open field test outcomes, wall colour of the open field arena and the reporting of measures to reduce the risk of bias were found to impact results. For in vitro electrophysiology experiments measuring synaptic plasticity, several electrophysiology parameters, including magnesium concentration of the recording solution, were found to explain a significant proportion of

the heterogeneity. Automated meta-research approaches and curated web platforms summarising preclinical research could have the potential to accelerate the conduct of systematic reviews and maximise the potential of existing evidence to inform translation.

LAY SUMMARY

Alzheimer's disease is characterised by severe memory impairment and a gradual loss of physical and mental capacity. Despite decades of intensive research, there is no cure for this detrimental condition. A crucial step in the development of new treatments involves testing promising new drugs in animal disease models. Tens of thousands of papers in scientific journals describe experiments modelling Alzheimer's disease symptoms and underlying brain changes in animals. However, treatments with promising results in these models often then fail to improve patient symptoms when later tested in clinical trials. By looking closely at the evidence from these models and evaluating it, we may be able to understand and overcome this problem. Systematic reviews are a method of collecting all available evidence on a given topic, evaluating it, and assessing the strength of the evidence. Taking things further, a meta-analysis is a statistical method which combines the results of lots of experiments. As meta-analyses aggregate the results from a greater number of people or animals, we can have greater confidence in the results. In healthcare settings, systematic reviews and meta-analyses are used frequently to decide, based on evidence from lots of experiments, whether a certain treatment is effective in most patients, or to assess whether something is harmful. These methods are less frequently used within laboratory based research in animal models as there is often much more variation between the results of animal experiments. However, we can use these methods to understand the sources of variation between studies and how differences in study design affect results. Furthermore, by summarising and assessing the quality of evidence in animal models, we can move into clinical trials with greater confidence. A key issue which limits the utility of systematic reviews of animal research is that with thousands of animal experiments conducted every year, it is near-impossible to keep up with all the latest evidence.

To address this problem, I have piloted the use of cutting-edge technology to build an automated system to collect, categorise, and display evidence from research papers on Alzheimer's disease animal models. I have also created publicly available, interactive online application to display all the information I have collected, which allows others to make use of it. Scientists can use this platform to assess the quality and the quantity of available evidence and to identify gaps where further research is required. Other stakeholders in Alzheimer's disease research (including research funders, institutions, patients, and their caregivers) can also use this platform to gain an overview of the current evidence.

I developed and validated new automated tools to extract details about the animal model(s), drug treatment(s) used, and the ways in which Alzheimer's related outcomes were measured from within publications. I also developed and validated a tool to identify and remove duplicates of the same publication within the dataset. To support automated evaluation of study quality, I built on existing work to determine whether a study incorporated steps to reduce the risk of unconscious bias (such as randomly allocating animals to experimental groups, or ensuring that experimenters do not know which animals are being treated with a drug throughout the study).

Using automated tools, I also identified two subsets of publications reporting experiments measuring (1) electrical activity between brain cells and (2) motor activity and anxiety levels in animal models of Alzheimer's disease. I used these datasets to perform more detailed systematic reviews of the literature in respect of which I summarised the existing evidence across commonly used Alzheimer's disease mouse models. Across each review, I identified certain aspects of experimental design which had an impact on results. Furthermore, I recognised that there were some differences between animal models and the human condition, such as animals typically being measured early in the lifespan before Alzheimer's disease has progressed. There was also a tendency to use male animals, with fewer

experiments measuring female mice. Additionally, experiments were often not reported in enough detail, which limited my ability to use all the data effectively in the meta-analysis. Based on findings from these reviews, I have pinpointed areas for research improvement and identified gaps in the literature where more research is needed.

CHAPTER 1: HOW CAN WE BEST UTILISE EVIDENCE FROM ANIMAL STUDIES TO DEVELOP SUCCESSFUL ALZHEIMER'S DISEASE THERAPIES?

1.1 Brief Introduction

"the greatest challenge for health and social care in the 21st Century"
(Livingston et al., 2017)

Alzheimer's disease (AD) is the most common cause of dementia, a progressive neurodegenerative disorder characterised by a severe loss of memory and cognitive capacity. AD places a huge burden on society, both emotionally and economically. Current therapeutic interventions aim to reduce the rate of disease progression, but do not significantly alter the underlying pathology (Agatonovic-Kustrin, Kettle, & Morton, 2018). The risk of developing AD rises sharply with age (Evans, 1990), leading to concerns about the impact this could have on an aging population. An estimated 5% of the European population are currently living with AD (Niu, Alvarez-Alvarez, Guillen-Grima, & Aguinaga-Ontoso, 2017), and prevalence is predicted to increase by 3 to 4-fold by 2050 (Rice et al., 2001). However, there is evidence that incidence within each age group is declining, perhaps due to improved lifestyle factors (Derby, Katz, Lipton, & Hall, 2017; Seblova et al., 2018). Although, it remains to be seen whether this trend can offset the impact of people living longer.

The major hallmarks of an AD brain are the presence of amyloid plaques consisting of aggregated amyloid beta (Aß) peptides, neurofibrillary tau tangles (hyperphosphorylated tau protein aggregates), neuroinflammation, synapse loss, and progressive widespread brain atrophy (Long & Holtzman, 2019; Serrano-Pozo, Frosch, Masliah, & Hyman, 2011).

1.2 The Amyloid Hypothesis

The so-called “amyloid hypothesis” is a widely held view that AD pathology is driven, first and foremost, by the accumulation of Aß in the brain (Hardy & Higgins, 1992; Selkoe, 2002). First proposed in 1991 (Hardy & Allsop, 1991), it was suggested that pathogenic alterations in the amyloid precursor protein (APP) gene led to abnormal Aß accumulation, beginning a cascade of events culminating in tau phosphorylation, tangle formation, and neuronal death.

APP is a transmembrane protein which has been extensively studied in relation to AD pathogenesis. APP is part of a family of highly conserved proteins, including amyloid precursor-like proteins (APLP1, APLP2) with homologs in mammals (Zheng & Koo, 2006), and amyloid precursor protein-like (APPL) proteins in fruit flies (Ewald & Li, 2012). APP differs in that its proteolysis can be amyloidogenic, leading to the generation of Amyloid beta (Aβ). In brief, cleavage of human APP by α-secretase and β-secretase generates membrane-attached carboxl terminal fragments (CTFs) and secretes soluble APP fragments (sAPPα, APPsβ). Sequential proteolytic cleavage of CTFs (namely C99) by neuronal β-site APP cleaving enzyme (BACE) and γ-secretase produces Aβ proteins (see Figure 1). The PSEN1 and PSEN2 genes, encoding for presenilin-1 and presenilin-2 are also critical to this pathway. The presenilins form part of the γ-secretase complex which cleaves APP to generate Aβ.

In this amyloidogenic cascade, Aβ40 (40-amino-acid), and the more toxic Aβ42 (42-amino-acid) proteins, are excreted into the extracellular space. In healthy individuals, Aβ is continually cleared into the peripheral circulatory or lymphatic systems. However, in AD, it has been hypothesised that a combination of defective clearance systems, Aβ overproduction, and/or conformational changes in amyloid species, lead to abnormal Aβ accumulation into neurotoxic plaques (Zuroff, Daley, Black, & Koronyo-Hamaoui, 2017).

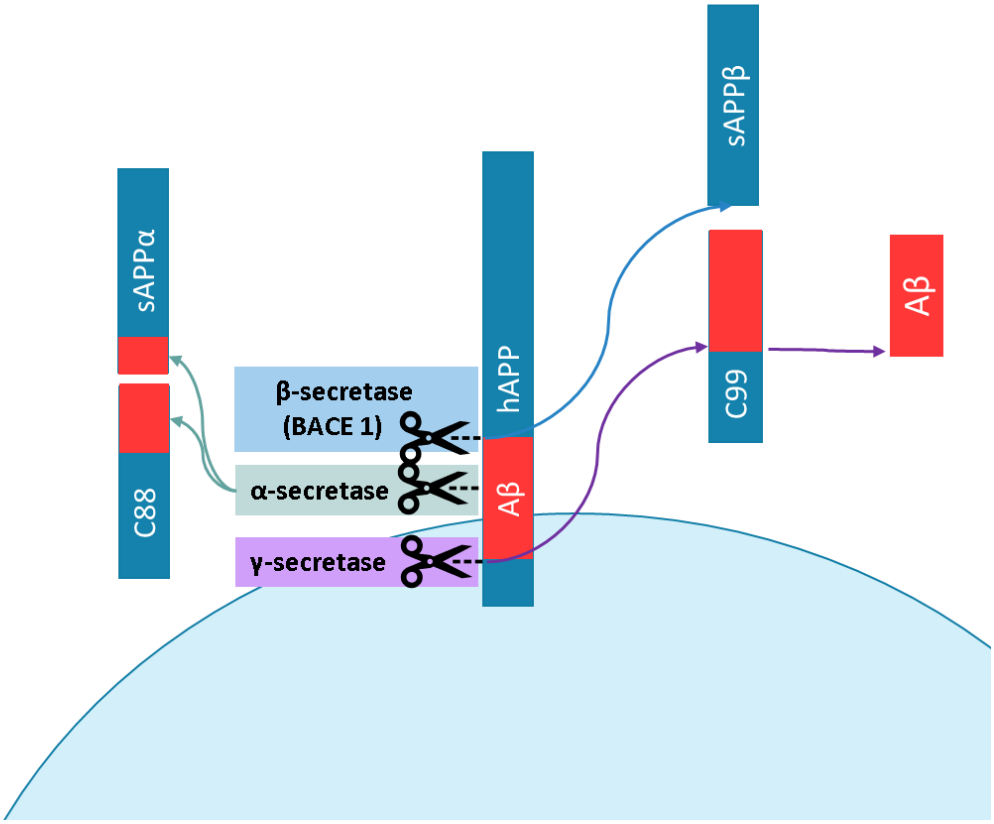


Figure 1.1: Simplified schematic of AB generation through cleavage of the APP gene

The theory that Aβ accumulation is crucial in the development of AD pathology (Hardy & Selkoe, 2002) is supported by a significant body of clinical evidence. Mutations in the APP gene encoding for the APP protein have been identified in families with early-onset familial AD (Goate et al., 1991; Murrell, Farlow, Ghetti, & Benson, 1991) and were among the first identified genetic markers of the disease. In fact, more than 100 different mutations across APP, PSEN1, and PSEN2 genes have been identified as contributing directly to familial AD (Nestler, Hyman, & Malenka, 2001). A notable APP mutation includes the double Swedish mutation, first identified in a large Swedish family (Mullan et al., 1992), where two base pair conversions were found in exons 16 and 17 of the APP gene near the site of β-secretase cleavage. The mutation leads to Aβ overproduction, and research has suggested that this mutation alone is sufficient to cause subsequent disease

pathology (Lannfelt et al., 1994). Mutations near the y-secretase site have been found to lead to increased production of Aβ42 (Chartier-Harlin et al., 1991; Goate et al., 1991), while other APP mutations affect the likelihood of Aß aggregation to form plaques (Tomiyama et al., 2008).

Further indication of the significance of alterations in amyloid processing comes from individuals with trisomy 21 (Down's syndrome), who carry one extra copy of the APP gene. These individuals develop amyloid plaques and neurofibrillary tangles from a young age and are at a significantly increased risk of early-onset Alzheimer's disease (Wiseman et al., 2015).

However, the idea that amyloid is the key to AD pathology has been challenged by findings that Aβ accumulation does not correlate with the severity of dementia symptoms (Dickson et al., 1995; McLean et al., 1999; Terry et al., 1991). Furthermore, until very recently, all therapeutic treatments targeting both Aß production (including B and y-secretase inhibitors) and clearance (via monoclonal antibodies which bind to and remove Aß plaques) have failed in clinical trials. However, this chain of successive failures may just have been broken by Aducanumab, an anti- Aß immunotherapy drug recently approved by the US Food and Drug Administration for use in AD. As the first new AD drug treatment to emerge in over 18 years, this development will no doubt spark a huge amount of hope and potentially re-ignite belief in the role of Aß accumulation as a key mechanism underlying AD. There remains a substantial amount of uncertainty about the true efficacy of Aducanumab and whether it can alter disease progression in a meaningful way (Alexander, Emerson, & Kesselheim, 2021; Mullard, 2021). Three members of the FDA resigned upon its approval, with one writing that to do so was "probably the worst drug approval decision in recent US history" (Mahase, 2021).

1.3 The Tau propagation Hypothesis

The other major hallmark of AD brains is the presence of intracellular neurofibrillary tangles (NFTs), composed of aggregated tau protein. Tau is encoded by the microtubule-associated protein tau (MAPT) gene and belongs to a family of microtubule-binding proteins. Under normal conditions, tau works to stabilise the structure of microtubule filaments which play an essential role in transporting proteins along axons (Barbier et al., 2019). Abnormal hyperphosphorylation of tau protein leads to conformational changes. Tau then self-aggregates to form bundles of paired helical filaments which become NFTs (Grundke-Iqbal et al., 1986). The formation of aggregated NFTs is thought to disrupt local communication between neurons, de-stabilise microtubules, and potentially propagate neurotoxic effects throughout the brain in a prion-like manner (Liu, Xie, Meng, & Kang, 2019).

Given the failure of amyloid targeting drug therapies, attention has turned in recent years to tau as an alternative, or complementary, drug target of interest. Therapeutic strategies targeting pathological tau have included immunotherapy approaches to clear tau, treatments to prevent aggregation (Gauthier et al., 2016), and treatments to prevent hyperphosphorylation of tau proteins (Li & Götz, 2017). However, to date, many of these approaches have not moved beyond pre-clinical *in vivo* testing. Of those which have made it to clinical trials, none have so far been successful in improving cognitive function in AD.

1.4 Other Hypotheses of AD Pathology

In this introductory chapter, I have detailed the mainstream hypotheses of AD pathology which are most relevant to my own work. However, as a complex and multifactorial disorder, researchers have proposed a huge array of possible underlying causes, both environmental and genetically pre-determined which may influence AD pathology. It has been recognised for decades that AD patients experience changes in the cerebral microvasculature (Buée et al., 1994). High

cholesterol diets and obesity, which may drive vascular changes, have been linked to AD incidence (Hassing et al., 2009; Proitsi et al., 2014). There is also evidence that neuroinflammation may play a role, mediated by an increase in microglia activation and inflammatory cytokine release (Santos, Beckman, & Ferreira, 2016). However, as for all proposed hypotheses of AD development, it remains difficult to separate which events trigger AD induced cognitive and structural decline and which pathological events are simply reactions to AD.

1.5 The Use of Transgenic AD Animal Models

In vivo models are useful tools to understand aspects of AD pathogenesis and to trial the effectiveness of drug therapies. Mouse models have been used most extensively across AD research for largely practical reasons; their entire genome has already been sequenced and their gestation period is short, simplifying breeding cycles. This has allowed the development of a multitude of transgenic AD models with familial AD mutations. Transgenic mouse models with mutations in the APP gene are most dominant across the preclinical literature and recapitulate some, but not all, aspects of AD pathology. Most develop an abundance of amyloid plaques and display changes in cognitive function. However, most models lack NFTs, synapse loss, and the widespread neuronal atrophy seen in the human condition.

A smaller number of experiments have used alternative approaches to induce sporadic AD-like phenotypes through injection of inflammatory or toxic compounds, dietary changes, or the direct application of Aß and tau proteins. Such approaches may be of more relevance to the sporadic form of AD (Lecanu & Papadopoulos, 2013; Shree, Bhardwaj, Kashish, & Deshmukh, 2017)

1.5.1 Early APP mutant models

As the genetic understanding of familial AD became known, researchers quickly developed mouse lines that overexpressed wild-type human APP (hAPP) (Buxbaum,

Christensen, Ruefli, Greengard, & Loring, 1993; Lamb et al., 1993). The resulting mice displayed mild neurobehavioral deficits and did not develop amyloid plaques. Just two years later, a transgenic mouse line (PDAPP) was successfully developed to overexpress a mutated form of hAPP (Games et al., 1995), which bore a closer resemblance to the amyloid pathology seen in the human condition. Since then, a variety of mouse models have been established which overexpress hAPP at several times endogenous levels, being necessary to achieve the desired pathology within the mouse lifespan. Many mutant APP models have been found to show an age-related cognitive decline on neurobehavioral outcomes and increasing amyloid plaque burden (Kitazawa, Medeiros, & M LaFerla, 2012). However, there remains a lack of clarity in how individual APP mutations contribute to the AD-like pathophysiology reported in different models.

1.5.2 Commonly used APP models in preclinical research

The PDAPP line was the first true AD mouse model with increased Aß load within the brain. These mice harbour the Indiana APP mutation and express humanised hAPP at ten times endogenous mouse APP levels (Games et al., 1995). Tg2576 mice carry the Swedish mutation and express mutant hAPP at approximately five-fold normal levels. In this model, diffuse Aß plaques develop at a slower pace, typically starting around 11-13 months (Hsiao et al., 1996). APP23 mice also carry the Swedish mutation, but the transgene is overexpressed at seven times normal levels as the mutation occurs on an alternative APP isoform and is under the control of a different promotor (Sturchler-Pierrat et al., 1997). Early reports suggested amyloid plaque deposition occurs as early as 6-months of age, but later evidence suggests the timeframe is much later at 15-18 months (Balducci et al., 2010). TgCRND8 mice carry both Swedish and Indiana mutations and develop Aß pathology at a much accelerated rate. In this model, hAPP is overexpressed at five times endogenous levels (Chishti et al., 2001). The J20 line also expresses both Swedish and Indiana

mutations, but on a different APP isoform (770) and under a different promotor (Mucke et al., 2000). Reports indicate that early indicators of Aß pathology begin at just 1 month of age (Hong et al., 2016), and progresses to widespread Aß plaques by 8-10 months of age.

Mutations in PSEN1 have also been introduced into mouse models to hasten pathology. These mutations shift APP processing by γ-secretase to produce longer, more pathogenic Aβ peptides which are useful for accelerating disease in short-lived animal models. APPswe/PSEN1dE9 mice harbour the Swedish APP mutation and a PSEN1 mutation lacking exon 9 and begin to show amyloid deposition by 6 months of age (Jankowsky et al., 2003). The 5xFAD model combines 5 different familial AD mutations across APP PSEN1 and exhibits rapidly progressing AD pathology. Importantly, for relevance to the human condition, this is accompanied by neuronal loss across multiple brain regions (Oakley et al., 2006).

The majority of APP and APP/PS1 models, including 5xFAD, do not show evidence of NFT development. To remedy this, in 2003, triple transgenic mice with an aggressive AD phenotype were generated with mutations in APP, PSEN1 and MAPT genes (Oddo et al., 2003). A summary of commonly used APP transgenic models are listed in Table 1.1.

	Transgene	**APP isoforms overexpressed**	**Overexpression details / amyloid pathology**
Single APP strains			
PDAPP (Games et al., 1995)	Human APP V717F (Indiana)	695, 751, 770	10x higher APP; AB pathology present at 6 months of age
Tg2576 (Hsiao et al., 1996)	Human APP KM670N/671NL (Swedish)	695	5x higher APP Aβ plaques present from 12 months of age
APPSwe (line C3-3) (Borchelt et al., 1996)	Chimeric mouse/human APP KM670N/671NL (Swedish)	695	3x higher APP; Aβ plaques present from 19 months of age
APP23 (Sturchler-Pierrat et al., 1997)	Mouse APP KM670N/671NL (Swedish)	751	7x higher APP; Aβ plaques present from 6 months of age

APPSweLon (Lamb et al., 1997)	Human APP KM670N/671NL (Swedish), APP V717I (London)	695, 751, 770	2-3x higher APP; increased AB42 generation at 3 months, but no plaque deposition
J20 (Mucke et al., 2000)	Human APP KM670N/671NL (Swedish), V717F (Indiana)	770	Aβ plaques present from 6 months of age
TASD41 (Rockenstein, Mallory, Mante, Sisk, & Masliaha, 2001)	Human APP KM670N/671NL (Swedish), APP V717I (London)	751	Aβ plaques present from 6 months of age
TgCRND8 (Chishti et al., 2001)	Human APP KM670N/671NL (Swedish), V717F (Indiana)	695	Aβ plaques present from 3 months of age
Tg-SwDI (Davis et al., 2004)	Human APP KM670N/671NL (Swedish), APP E693Q (Dutch), APP D694N (Iowa)	770	Aβ plaques present from 3 months of age
Double APP/PS1 strains			
APPSwe/PSEN1dE9 (Jankowsky et al., 2003)	Mouse / human APP $_{M670N/671NL}$ (Swedish) Human PSEN1$_{deltaE9}$	695	3x higher APP; Aβ plaques present from 6 months of age
APPSwe/PSEN1(A246E) (Borchelt et al., 1996)	Mouse / human APP $_{KM670N/671NL}$ (Swedish) Human PSEN1 $_{A246E}$	695	Aβ plaques present from 9 months of age
APPPS1 (Radde et al., 2006)	Human APP KM670N/671NL (Swedish) Mouse PS1 $_{L166P}$	751	3x higher APP; Aβ plaques present from 1.5 months of age
PS/APP (Holcomb et al., 1998)	Human APP KM670N/671NL (Swedish) PSEN1$_{M146L}$	695	Aβ plaques present from 6 months of age
5xFAD (Oakley et al., 2006)	Human APP KM670N/671NL (Swedish) I716V (Florida), 717I (London) PSEN1$_{M146L, L286V}$	695	3x higher APP; Aβ plaques present from 3 months of age
Triple APP/PSEN1/MAPT strains			
3xTg-AD (Oddo et al., 2003)	Human APP KM670N/671NL (Swe) PSEN1$_{M146V}$ MAPT$_{P301L}$	695	Aβ plaques present from 6 months of age

Table 1.1: Commonly used APP mouse models.
Three major isoforms of APP exist (APP 695, 751, and 770), produced by alternative splicing of the APP gene. The mutated isoforms overexpressed by each transgenic model are shown in the third column. Details of amyloid pathology and overexpression of mutated APP isoforms (versus endogenous APP) are shown in column four.
Summarised from a range of review articles (Jankowsky & Zheng, 2017; Myers & McGonigle, 2019; Sasaguri et al., 2017) and Alzforum.org.

1.6 Translational failure in AD research

Billions of pounds have been spent on research conducted across pharmaceutical industries and academic institutions in a concerted effort to develop disease-modifying treatments for AD. Positive efficacy results from preclinical trials in animal models have encouraged numerous high-profile and highly anticipated clinical trials of new compounds. However, nearly all have ended in bitter disappointment. Of those drug candidates which reach human clinical trials, 99.6% have ultimately not resulted in any meaningful improvements in AD symptoms or progression (Cummings, Morstorf, & Zhong, 2014). Since the start of the millennium, over 400 clinical trials testing AD treatments have failed (Rinaldi, 2018).

Faced with this bleak outlook, we must consider whether our reliance on *in vivo* animal studies is misplaced, or whether clinical trials are conducted inappropriately? This thesis concentrates on the former. I will focus my attention on the reliability, transparency, and quality of preclinical evidence, and how we can synthesise and critically appraise it to guide improvements in translational potential.

1.6.1 How reliable are animal studies?

Translation from bench to bedside rests upon reliable evidence from well-designed preclinical studies. However, there is a growing awareness that we are unable to replicate the majority of preclinical research findings (Begley & Ioannidis, 2015; Peers, Ceuppens, & Harbron, 2012), and results can be highly variable even within a single animal model of neurological disease. Widespread methodological flaws across the preclinical literature, coupled with a lack of transparency about how experiments were conducted, (Bahor et al., 2017; Egan & Macleod, 2014; Macleod et al., 2015; Veening-Griffioen et al., 2019) contribute to and exacerbate the issue.

The confidence we have in any given experiment relies upon the internal validity – or the extent to which an experiment can support a claim about a cause-and-effect relationship. Internal validity is threatened by biases which can arise at different stages of a study (de Vries et al., 2014), including systematic differences in baseline characteristics of experimental and control groups (selection bias), in animal care, handling, or treatment administration between groups (performance bias), in outcome assessment between groups (detection bias), and the way in which animal and datapoint exclusions are determined (attrition bias). Persistent failures to report measures that reduce these risks of bias across the biomedical literature have been associated with inflated estimates of treatment efficacy and likely lead to false positive results, where a drug appears to improve outcome but in reality does not (Bello et al., 2014; Crossley et al., 2008; Hirst et al., 2014; Tsilidis et al., 2013). Even if true positive (real) treatment effects are present, underpowered experiments may fail to detect them. Small sample sizes (generally <10 animals per group) are commonplace in preclinical research, but when an experiment is not adequately powered, the confidence placed in the results decreases while the variability of results may increase. As demonstrated in the field of Amyloid Lateral Sclerosis (Scott et al., 2008), underpowered studies can severely reduce the usefulness of preclinical evidence and hinder translation. Although likely to be of similar importance in AD research, a SR review dataset of over 400 studies testing interventions in animal models of AD (Egan & Macleod, 2014), identified no studies reporting a power calculation to determine sample size.

The Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines were introduced in 2010 (Kilkenny, Browne, Cuthill, Emerson, & Altman, 2010) with a recent update in 2020 (Percie du Sert et al., 2020). The guidelines were introduced as a quality item checklist for authors - with the intention of improving the transparency of research using animals. Unfortunately, there is no evidence it has led to significant improvements in reporting quality across AD research (Egan &

Macleod, 2014; Veening-Griffioen et al., 2019), or across the wider biomedical landscape (Avey et al., 2016; Hair, Bahor, Macleod, & Sena, 2020; Kilkenny et al., 2010). It is important to consider that poor reporting does not immediately imply poor conduct – but survey evidence from *in vivo* researchers suggests that the conduct of experiments *was* in line with what is reported or, in some cases, was less rigorous (Reichlin, Vogt, & Würbel, 2016). In either case, without methodological clarity, we cannot efficiently replicate studies nor easily determine the value of available evidence.

Given that "positive" results in preclinical studies inform drug candidate selection, these weaknesses should not be underappreciated. This is especially true when considering the ethical concerns and potential harms to patients who are given drug treatments that have not been adequately evaluated for their likelihood of benefit.

1.6.2 How applicable are rodent models to the human condition?

When designing preclinical studies, the external validity – the extent to which findings in *in vivo* models can generalise across different settings – must also be considered. We can evaluate the validity of a disease model using three primary criteria: face validity (i.e. the similarity between the human condition and the model phenotype), predictive validity (i.e. the similarity between treatment effects observed in the model and treatment effects in the human condition), and construct validity (i.e. the similarity between theoretical or confirmed disease mechanism in the model and the disease mechanism in humans) (McKinney & Bunney, 1969; Sams-Dodd, 2006; Willner, 1984).

Construct validity is perhaps the most difficult to assess, as although we have theoretical frameworks (e.g. the amyloid hypothesis, synaptic failure, ageing, microglia/neuroinflammation), the underlying mechanisms of AD are not clearly defined. A small proportion of AD cases arise due to rare autosomal dominant

mutations in genes encoding for the amyloid precursor protein and presenilin (Bertram, Lill, & Tanzi, 2010). Although inherited, or "familial", AD accounts for less than 5% of AD patients (Wu et al., 2012), these mutations were the starting point for developing the genetically modified disease models which are most prominent in preclinical studies. The first transgenic (Tg) mouse line attempting to mimic AD pathology was created over two decades ago (Games et al., 1995) and since then at least 160 transgenic rodent models have been developed (Cuello, Hall, & Do Carmo, 2019; Mullane & Williams, 2019).

Rodents do not naturally develop AD, and so the pathology induced in transgenic models can never be fully representative of the human condition. Instead, all models can be seen as reductionist recapitulations of AD which has fuelled long-standing concerns around face validity (Franco & Cedazo-Minguez, 2014). The defining features of AD pathology include widespread brain atrophy, amyloid plaques (consisting of aggregated amyloid beta (Aß) peptide), neurofibrillary tau tangles and synaptic loss. However, as these hallmarks often cannot be reliably detected in living patients, the key endpoint of most clinical trials is cognitive decline (Andrieu, Coley, Lovestone, Aisen, & Vellas, 2015). One concern is that the rodent models we test may not have the intellectual capacity to evaluate this. Rats have a more complex brain structure and superior cognitive abilities (Zimmer, Parent, Cuello, Gauthier, & Rosa-Neto, 2014), and may have a greater potential for testing complex behaviours (Do Carmo & Cuello, 2013; Pressler & Auvin, 2013). Critically, most animal models do not exhibit the large- scale synaptic loss which occurs in AD. Unlike amyloid plaque density, this aspect of AD pathology is highly correlated with cognitive impairment (Robinson et al., 2014; Scheff, Price, Schmitt, & Mufson, 2006; Terry et al., 1991). However, many animal models do display deficits in synaptic plasticity and transmission (Marchetti & Marie, 2011) and therapeutic potential may lie in targeting synaptic dysfunction at early stages of the

disease (Koffie, Hyman, & Spires-Jones, 2011; Nistico, Pignatelli, Piccinin, Mercuri, & Collingridge, 2012).

A recent study (Veening-Griffioen et al., 2019) aimed to quantitatively assess predictive validity across a variety of AD models by comparing cognitive outcomes for interventions tested in preclinical studies and clinical trials. Veening-Griffo and colleagues reviewed the literature and identified major shortcomings in the external and internal validity of animal studies. Citing these factors, they were not surprised to find that the predictive value of any model was divergent across interventions and that individual interventions had different efficacies across models. This led the authors to conclude that "currently, no animal models can be recommended for determining the efficacy of interventions in Alzheimer's disease". According to large SRs of the preclinical AD literature, interventions are, on average, administered early in the rodent lifespan (Egan & Macleod, 2014), in some cases months before evidence of plaque burden and cognitive impairment emerges. While preventative treatments may be beneficial, to have any likelihood of translation they must inform clinical trials which begin prior to disease onset in those most likely to develop AD. Furthermore, testing interventions so early in disease progression exacerbates the disparities in disease duration experienced by rodents versus patients – if it takes AD patients a decade for the neuropathological features of the disease to progress enough to cause symptoms and rodents only a couple of months, it is difficult to imagine how a treatment given in the former would compare to the latter.

AD is a complex, multifactorial disorder. While no AD model can be fully representative of the human condition, some may be more relevant for investigating individual aspects of the disease (e.g. tau pathology) than others (Shineman et al., 2011), and could provide further mechanistic insights into how these facets of the disease manifest and the relative strengths and weaknesses of each model (McGonigle & Ruggeri, 2014). As pointed out by others (Quinn, 2018),

new therapies have been identified over the last decade to treat multiple sclerosis (MS) – another highly complex neurological disease – which indicates that lack of a "perfect" animal model doesn't prevent progress. In MS, there have been efforts to selectively target the inflammatory aspects of the disease which are reproduced in animal models (Constantinescu, Farooqi, O'Brien, & Gran, 2011), and successful clinical trials have appropriately aligned their outcomes with the preclinical evidence. Furthermore, to better evaluate construct validity, we need to understand how facets of the disease relate to each other, i.e. is cognitive decline in the animal model also correlated with synaptic loss / dysfunction? If a drug only improves amyloid plaque load and moderately improves performance on cognitive tests, is there enough likelihood of benefit to trial on AD patients? Given the concerns surrounding internal and external validity, it is evident that no preclinical study using AD mouse models can be used in isolation to justify the selection of compounds for clinical trials.

1.7 Making sense of the preclinical evidence

"We can't solve problems by using the same kind of thinking we used when we created them".

- Albert Einstein

Instead of continuously investing in new drug targets in the hope that the next will be successful, the past decade has taught us that we need to re-examine how we arrived here and re-evaluate our next steps.

Some have called for a radical rethink of preclinical research and its contribution to medicine; even going as far as to say that we should stop and take stock before conducting any more animal experiments (Pound, Ebrahim, Sandercock, Bracken, & Roberts, 2004). Nevertheless, after billions of pounds, millions of animals, and thousands of experiments, should we just stop what we are doing and start again?

One approach is to utilise the existing evidence we have by conducting a wide-ranging systematic review (SR) of the literature. SRs allow researchers to gain an overview of the evidence, determine the quality of that evidence, and identify which experimental design factors may influence the reproducibility and predictive value of animal studies (de Vries et al., 2014). Furthermore, by pooling the results from included studies in a meta-analysis, we can attain an overall effect size for different experimental interventions and attempt to explain some of the heterogeneity between the studies.

1.7.1 A brief history of evidence synthesis

Scientists have recognised the need to effectively synthesise research evidence for hundreds of years (Chalmers, Hedges, & Cooper, 2002). However, SR methodology did not emerge until the latter half of the 20th Century, when proposed as a tool to enable clinicians to make evidence-based decisions about patient care. SRs are performed regularly in the clinical domain and the issues highlighted have been used to inform improvements in clinical trial design (Mullen & Ramírez, 2006).

In contrast, preclinical SRs are a relatively new approach. After a highly anticipated novel stroke intervention, Nimodipine, failed to deliver any benefits in human clinical trials, the quality of preclinical evidence that had supported the transition to human use was called into question (Horn, de Haan, Vermeulen, Luiten, & Limburg, 2001). In the backlash that ensued, a landmark commentary proposed that a SR of prior evidence from relevant animal experiments should form an essential part clinical trial design (Sandercock & Roberts, 2002). As discussed, there are many challenges which may impede the translation of findings from animal models to human patients. SRs of animal studies can help expose the reasons behind this discrepancy (Hooijmans & Ritskes-Hoitinga, 2013).

In a retrospective review of the preclinical evidence that informed 6 high-profile AD drug candidates that failed in clinical trial (Karran & Hardy, 2014), the authors concluded that some were "very unlikely to succeed" based on the evidence. Of those reviewed, 4 (Tramiprisate, Semagacestat, Bapineuzumab, Solanezumab) had incomplete or inconsistent *in vivo* data from animal models and 2 had *in vivo* data which did not support progression to phase 1 clinical trials (Tarenflurbil, Gammagard). Although some of the pitfalls of the compounds were known at the time, a thorough and rigorously conducted systematic review of the evidence could have provided clear guidance about where the gaps were, how strong the evidence was for a specific outcome to be measured in patients (e.g. the ability to reduce levels of existing amyloid plaques OR cognitive improvements), and the likelihood of benefit.

1.7.2 A systematic approach to evidence collection and curation

"[Scientists] are little better than laymen at characterizing the established basis of their field."

- Thomas Khun

SRs are intended to be of high methodological quality and avoid biases which can skew conclusions in traditional literature reviews (Oxman & Guyatt, 1993). A SR should identify and critically appraise all available evidence for a given research question. Firstly, it enables researchers to obtain a less biased overview of the evidence than would otherwise be possible. Thomas Kuhn once stated that scientists may "*learn easily and well about the particular individual hypotheses that underlie a concrete piece of current research*," but despite this, "*they are little better than laymen at characterizing the established basis of their field*." This is perhaps true across all research domains, although preclinical research relevant to human health poses a particularly striking example. Even if one attempted to keep up to date, the sheer volume of (often conflicting) experimental findings may be

insurmountable. In narrative reviews, an expert author selects studies to discuss which they feel are of relevance, however, this methodology is neither transparent nor reproducible. Furthermore, although no sound evidence exists for preclinical studies, citation bias – or the tendency to cite studies finding statistically significant effects – plagues medical research and can inflate the perceived effectiveness of treatments (de Vries et al., 2018; Jannot, Agoritsas, Gayet-Ageron, & Perneger, 2013).

Using SR approaches, search strategies are developed to systematically obtain relevant records from databases. The title and abstract (or full-text) of retrieved publications are screened according to strict inclusion criteria and prespecified, structured information from included publications is then extracted. To take things further, summary data from the publication (displayed in tables or figures) can be used to calculate effect sizes with precision estimates. The results of similar included studies can be pooled in a meta-analysis in which calculates an overall summary estimate of effect for a group of studies, with more weight given to more precise studies. Differences between summary effects, whether between groups of studies testing different drugs or studies which reported randomisation versus those which did not, can be assessed. This extent of this difference, or heterogeneity, can be explored to identify which aspects of the experimental design have the biggest impact on effect sizes we observe. The steps involved in conducting a preclinical SR and meta-analysis are shown in Figure 2.

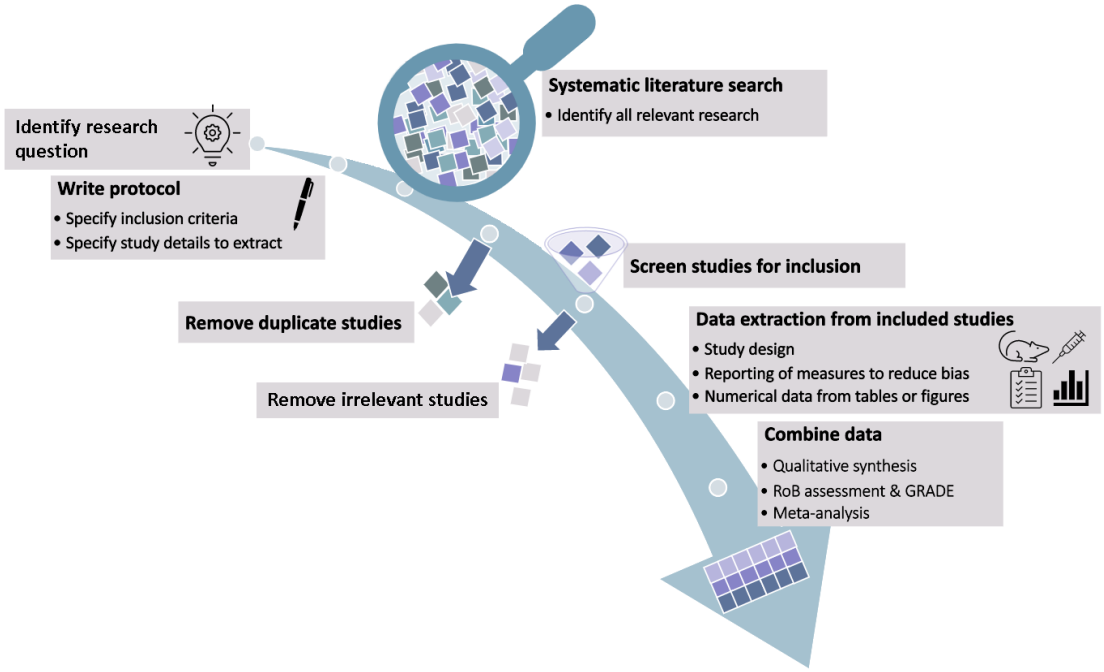


Figure 1.2: Steps involved in preclinical SR and meta-analysis

1.7.3 Using SR evidence to guide future research

Evidence synthesized by clinical SRs is viewed as gold-standard, however, due to the variability and validity of preclinical research findings, they are primarily hypothesis-generating. In contrast to clinical research, preclinical studies have smaller samples sizes and much greater heterogeneity, so the objective is often to explore and understand sources of variation, with aims to improve the rigour and transparency of future studies (Sena, Currie, McCann, Macleod, & Howells, 2014).

Even with this caveat, preclinical SRs can still inform decision making. Their findings can help determine where the gaps are in each research area, whether there is enough good quality preclinical evidence for a compound to proceed to clinical trial, and which areas of study design or reporting should we prioritise for improvement. SRs accompanied by a meta-analysis can also inform which aspects of experimental

design may need to be controlled or systematically varied to have more confidence in a biological phenomenon or treatment effect.

Depending on the research question, SR findings can describe which interventions have been tested on which disease model, highlight differences in methodology, and pinpoint gaps in the literature. Past reviews of the preclinical literature have highlighted persistent failures to report measures to reduce the risk of bias such as randomisation, allocation concealment, and blinded outcome assessment (Bahor et al., 2017; Egan & Macleod, 2014; Macleod et al., 2015), and poor reporting of quality items recommended ARRIVE guidelines (Avey et al., 2016; Kilkenny et al., 2010). Furthermore, the external validity of preclinical studies can be compromised if significant differences exist between animal disease models and the human condition. For example, in the AD literature, primarily male mice are used despite AD disproportionately affecting female patients (Ferretti et al., 2018) and most models recapitulate only a few individual aspects of the human condition (Franco & Cedazo-Minguez, 2014) with the vast majority lacking the neuronal death predominantly seen in AD brain tissue.

Taking things further, summary data from the publication (displayed in tables or figures) can be used to calculate effect sizes with precision estimates. The results of similar included studies can be pooled in a meta-analysis which calculates an overall summary estimate of effect for a group of studies, with more weight given to more precise studies. Differences between summary effects, whether between groups of studies testing different drugs or studies which reported randomisation versus those which did not, can be assessed. The extent of this difference, or heterogeneity, can be explored to identify which aspects of the experimental design have the biggest impact on effect sizes we observe. In contrast to clinical research, preclinical studies have smaller samples sizes and much greater heterogeneity, so the objective is often to explore and understand sources of variation, with aims to improve the rigour and transparency of future studies (Sena et al., 2014). The

omission to report measures to reduce the risk of experimental bias have been found in meta-analyses to explain a significant proportion of the heterogeneity between studies and have been associated with inflated estimates of effect size (Bello et al., 2014; Crossley et al., 2008; Hirst et al., 2014).

Taken together, preclinical SRs and meta-analyses aim to provide an overview of the best available evidence. They can be used as tools to summarise what we truly know about a given field, benchmark the quality of reporting, and provide insights to guide and improve the design and conduct of future research.

1.7.4 Preclinical SRs: barriers to impact

The number of preclinical SRs is increasing (van Luijk et al., 2014), and several have already been published focussed on Alzheimer's models (Egan, Vesterinen, Beglopoulos, Sena, & Macleod, 2016; Ekert, Gould, Reynolds, & Howard, 2018; Foley, Ammar, Lee, & Mitchell, 2015; Hooijmans, Pasker-de Jong, de Vries, & Ritskes-Hoitinga, 2012; Yamasaki et al., 2012). However, there are several barriers that can reduce the translational potential and impact of SR findings.

(1) SRs should capture all available evidence, but this is not always the case (Bashir, Surian, & Dunn, 2018; Bastian, Glasziou, & Chalmers, 2010; Créquit, Trinquart, Yavchitz, & Ravaud, 2016). Inadequate search strategies, narrow research questions, and failures to incorporate new findings can hinder attempts to accurately synthesise and evaluate evidence. Previous SRs of the literature have identified between 899 (Veening-Griffioen et al., 2019) and 8,360 (Egan et al., 2016) relevant publications in animal models of AD, before excluding references based on predefined exclusion criteria e.g. studies not reporting primary experiments or not measuring a specific outcome.

(2) By the time of publication, SRs are frequently years out of date (Tricco, Brehaut, Chen, & Moher, 2008). Performing each stage of the SR process manually – searching of biomedical databases, removal of duplicate records, screening for relevance, and extracting information from included publications – is laborious and time-consuming (Thomas et al., 2017). Furthermore, updating SRs to incorporate recent evidence is not standard practice, and can require just as much effort as conducting the initial review (Lefebvre, Glanville, Wieland, Coles, & Weightman, 2013; Shojania, Sampson, Ansari, Ji, Garritty, et al., 2007)

(3) Without co-ordinated efforts to conduct and prioritise which SRs to perform, the degree of overlap can be significant (Siontis, Hernandez-Boussard, & Ioannidis, 2013), with little acknowledgement of findings from previous reviews on the same or similar research questions (Helfer et al., 2015).

(4) Sometimes the evidence SRs needs simply isn't available. Without unified standards around open-access publication, researchers are limited by their institutional subscriptions determining what they can and cannot access. Furthermore, the current publication format, most often presented with no raw data available, prevents reviewers from extracting outcome data directly or programmatically. Instead, researchers resort to the less accurate and more painstaking process of measuring the height of figures to get means and measurements of error.

(5) Publication-bias hides neutral and "negative" results. The "file drawer problem" of experiments which go unpublished introduces an unwanted bias into the SR process. Although not a solution, we can use statistical techniques such as Egger regression and trim-and-full analysis to estimate the extent of publication bias in a meta-analysis and try to understand the potential impact on results. In the preclinical stroke literature, publication

bias favouring positive results was found to account for an overestimation stroke treatment efficacy by around one third (Sena, van der Worp, Bath, Howells, & Macleod, 2010).

1.8 Aims and objectives

My work aims to provide a comprehensive summary of existing preclinical AD research, with a particular focus on frequently used and clinically relevant outcome measures in commonly used transgenic AD models. To accomplish this aim, throughout this thesis I have explored crowdsourcing and automation approaches to reduce the time taken to conduct SRs within this field.

In Chapter 2, I describe my work to develop and implement an automated evidence workflow and interactive online dashboard to summarise preclinical AD research in transgenic models. This workflow was made possible by crowdsourcing approaches to quickly obtain a dataset which could be used to train a machine learning algorithm to screen studies for relevance. Further, to make it easier to identify studies of relevance for specific SR questions, I developed text-mining dictionaries to classify studies by model, treatment, and outcome measure. I also built upon existing automated tools to assess risk of bias reporting.

In Chapter 3, I evaluate the performance of the text-mining tools I developed to enable this automated evidence workflow and discuss their current limitations.

In Chapter 4, I describe my work to develop and validate a novel, automated deduplication tool for removing duplicate records in systematic database searches. The removal of duplicate records in SRs is an under-appreciated problem. For larger scale reviews and evidence synthesis projects, as described in Chapter 2, substantial portions of the dataset will be duplicated, and a lack of user-friendly and scalable tools exist to tackle this issue, particularly within the preclinical domain.

In Chapter 5, using text-mining approaches, I obtained a subset of studies from my larger AD dataset (Chapter 2) to conduct a SR and meta-analysis of studies measuring synaptic plasticity in hippocampal slices derived from APP transgenic models. While synaptic loss is often not apparent in transgenic mouse models, there are mixed reports of synaptic dysfunction. In this review, I aimed to summarise the evidence across different AD models and understand the influence of different experimental protocols and reporting of indicators of study quality on results. I also aimed to integrate evidence from studies which also measured cognitive performance in the Morris water maze to understand the relationship between electrophysiological and behavioural outcomes.

Using a similar text-mining approach, I obtained a second subset of the AD dataset to conduct a SR and meta-analysis of studies in transgenic APP models measuring locomotor activity and anxiety in the open field test (Chapter 6). This is a commonly used behavioural measure across biomedical research, but there have been many reports of inconsistent findings and difficulty interpreting the biological meaning of different test measures. I aimed to critically evaluate the literature, explore reasons for the between study heterogeneity, and characterise the behaviour profile of different APP models.

In Chapter 7, I describe the application of automated evidence synthesis workflows, described previously (Chapter 2), to a novel research area, in response to the COVID-19 pandemic.

Finally, in Chapter 8, I summarise my findings from across the thesis chapters and discuss their relevance in the context of AD. I also discuss the feasibility of automated approaches in preclinical AD meta-research and the application of these technologies in a wider context.

CHAPTER 2: BUILDING AN AUTOMATED WORKFLOW TO SYNTHESISE THE ALZHEIMER'S DISEASE LITERATURE

2.1 Chapter Introduction

This chapter describes my work to develop, combine, and implement automated meta-research tools to summarise the preclinical AD literature, focussing on research conducted in transgenic AD models. This methodology builds upon efforts within our research group – Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies (CAMARADES) – to automate steps of the SR process. In this project, I aimed to obtain a high-level overview of the AD literature and the transgenic models used, interventions tested, and outcomes assessed – which would not be possible using traditional meta-research techniques. This workflow also enables SRs to be carried out at a much-accelerated pace, with less manual effort required in the initial stages. Methodology developed to enable this workflow is described in Chapters 3 and 4, while the SR projects enabled by this workflow are described in Chapters 5 and 6.

The annotated data used to train the machine learning algorithm described in this chapter was obtained with the help of many contributors (see Acknowledgements).

2.2 Background

2.2.1 Traditional SR methodologies are laborious

As discussed in Chapter 1, a major barrier which precludes preclinical SRs from having maximal impact is that they are labour-intensive and time-consuming (Borah, Brown, Capers, & Kaiser, 2017; Tricco et al., 2008). Evidence can be several years out of date by the time a review is complete; missing newer studies which may alter their conclusions (Shojania, Sampson, Ansari, Ji, Doucette, et al., 2007).

Incorporating new studies is an ongoing challenge with the rate of publication increasing exponentially each year (Bornmann & Mutz, 2015). For research areas

with a high rate of publication, the median ‘‘survival’’ of reviews (i.e., period over which they remain up to date) was found to be just under 3 years (Shojania, Sampson, Ansari, Ji, Doucette, et al., 2007). Additionally, SRs are not updated regularly or by following established approaches (Garner et al., 2016). In fact, so much work is often required to update a review that teams have described it as “just like starting a review from scratch” (Elliott et al., 2017). The resources required to perform and maintain SRs, reduces their reliability and utility.

2.2.2 Accelerating the speed of SRs

Recruiting a range of collaborators may assist in accelerating the pace at which a SR can be performed. Crowdsourcing, where a large group of external reviewers are recruited and trained to perform specific SR tasks, is a growing movement in meta-research. A recent evaluation concluded that a crowd of volunteer reviewers could accurately identify randomised controlled trials versus other types of research, without a substantial requirement for an expert reviewer to step in (Noel-Storr et al., 2017). Similarly, our research group have successfully employed a crowdsourced approach to assess over 700 publications against an ARRIVE checklist (Hair et al., 2019) and over 800 publications against the Nature journals’ reporting checklist (The NPQIP Collaborative Group, MacLeod, Sena, & Howells, 2019). A crowdsourced approach also offers the opportunity to expand the reach of SRs and educate more researchers about SR methodology. Those who are new to meta-research will likely encounter the same difficulties faced by more experienced meta-researchers in understanding publications in enough detail to extract relevant information. Greater awareness of these issues could facilitate an increase in the adoption of transparent and open research practices.

In recent years, our research group has also taken steps to facilitate faster research synthesis through automation technologies. To regularly retrieve new publications from PubMed, we have implemented auto-retrieval via application programming

interfaces (APIs) and enabled this as part of an online SR platform (SyRF) (Bahor et al., 2021). Simple, repetitive tasks are typically easier to automate than other more complex processes. Furthermore, crowdsourced approaches can help enable the development of automated tools by allowing us to quickly gather large amounts of useful training data (human decisions and annotations) on a subset of relevant research publications. For example, to accelerate the pace of citation screening, we have demonstrated the feasibility of training machine learning (ML) algorithms based on human screening decisions studies describing animal models of depression (Bannach-Brown et al., 2019), and animal models of neuropathic pain (Currie et al., 2019). We have also developed innovative text-mining techniques (see Chapter 3) to automate risk of bias assessment for preclinical studies (Bahor et al., 2017). Bringing together a range of techniques to increase the speed of SRs, the concept of a "living" SR (Elliott et al., 2014) – a systematic and continually updated summary of a research field, incorporating new findings as and when they emerge – has been proposed. However, some human effort will likely always be required to understand and extract detailed information and numerical data from publications. A summary of contemporary approaches to accelerate each step of the SR process is shown in Table 2.1.

Review Stage	Task	Potential methods
Team formation	Use a wider range of personnel than may traditionally be the case	• Crowdsourcing (e.g. Cochrane Crowd) • Task-sharing platforms (e.g. Task Exchange) • Webpage advertising projects linked to SyRF / part of living evidence summary
Search & citation download	Running search on multiple biomedical databases	• Automatic continuous database search push notification / RSS feeds • Notification from preclinical trial registries
	Obtaining metadata from records not included in citation	• Automatic retrieval of metadata from papers (e.g. Crossref, Endnote)
	Removal of duplicate records in dataset	• Automated and manual duplicate removal via Endnote, Zotero, Systematic Review Assistant-Deduplication Module (Rathbone, Carter, Hoffmann, & Glasziou, 2015), CAMARADES, Automated Systematic Search Deduplication tool (ASySD) available at: https://camarades.shinyapps.io/RDedup
Full-text retrieval	Downloading the full paper as a PDF, .txt or XML file for human and tool evaluation	• Automatic retrieval of full-text papers (e.g. Crossref, Endnote)
Eligibility assessment	Selecting studies for inclusion	• Machine-learning classifier (e.g. EPPI Centre) • Crowdsourced inclusion decisions
Data extraction	Extracting information on characteristics of the participants (animal models), interventions, and outcomes	• Regular expressions to detect model, intervention, and outcome level information
	Assessment of methodological rigor and risk of bias	• Regular expressions to detect risk of bias reporting • Machine-learning for automated risk of bias assessment (e.g. Robot Reviewer)
	Conducting meta-analysis	• Continuous analysis updating based on availability of structured extracted data • Meta-analysis web application for preclinical studies: https://camarades.shinyapps.io/meta-analysis-app-syrf/
Synthesis	Report writing and updating conclusions	• Templated reporting of some report items • Statistical surveillance of key analysis results, with threshold set for potential conclusion change
	Visualisation and interrogation of large datasets	• Interactive web applications: https://camarades.shinyapps.io/Preclinical-Models-of-Depression/ https://camarades.shinyapps.io/LivingEvidence_AD/

Table 2.1: Methods for efficiency gain at different steps of the SR process Adapted from (Thomas et al., 2017)

Within our group, we have been working towards a novel approach to summarise vast amounts of evidence from a chosen field of preclinical research. A recent SR of animal models of depression used text-mining techniques to automatically extract key methodological details from publications including the type of animal model used and the pharmacological interventions tested within a publication. The categorised dataset of over 18,000 publications was then displayed in an interactive application for others to use (Bannach-Brown, 2018). This approach was the first example of what we now term "systematic online living evidence summaries" or SOLES projects, which enhance the speed at which SRs can be performed and have the potential to transform the way we make sense of, and evaluate, research.

2.2.3 Systematic online living evidence summaries

In this chapter, I describe the SOLES approach of synthesising the evidence from preclinical animal studies within a specific research domain. SOLES are systematically obtained, up to date, curated, and interrogatable datasets of existing evidence. This is likely of most benefit to research-intensive areas where there is a high rate of publications. Given the high failure rate of Alzheimer's disease (AD) clinical trials, a SOLES for preclinical studies which model AD phenotypes and test novel compounds may be of significant value.

Much of the SOLES approach is borrowed from "living" SRs and utilises automation approaches to continually synthesise vast amounts of data. The major difference being scale. A typical SRs has a narrow, focussed research question while a SOLES encompasses an entire field. SOLES begin with a wider search strategy across several biomedical databases and a less stringent screening process – for example, including all primary experiments in any animal model of Alzheimer's disease. Some data is then extracted to categorise the research – for example, by disease model(s), intervention(s), and outcome measure(s). In addition, some quality measures may be extracted. This curated dataset can then be made accessible to

the research community via an interactive web application. This enables researchers to quickly ascertain the quantity and quality of research evidence, both overall and within each subgroup of studies employing each model, testing a certain treatment, or measuring a specific outcome. Furthermore, if the SOLES dataset sufficiently captures the existing literature, those wishing to perform a SR within that research domain can download a filtered set of references as a starting point, eliminating one of the most time-consuming tasks in the SR process (Bastian et al., 2010).

This approach integrates automated processes to benefit end users. Despite the existence of automation tools, there are significant practical and knowledge barriers which reduce their uptake by individual SRs teams (Al-Zubidy, Carver, Hale, & Hassler, 2017). Many tools exist as isolated use-cases which lack integration with other parts of the SR workflow and are may not be immediately available to researchers lacking the sufficient technical expertise (van Altena, Spijker, & Olabarriaga, 2019). For independent research groups working on similar SR projects, there may be immediate benefits in obtaining one single evidence base, through a concentrated and collaborated effort, upon which to derive all further, more detailed, reviews. In this chapter, I will describe my work to develop a prototype SOLES for preclinical AD research, utilising automation and crowdsourcing methodologies.

2.2.4 Aims of AD-SOLES

I aimed to develop, validate, and pilot an innovative evidence summary that has the potential to provide a continually updated synthesis and assessment of preclinical AD research conducted in transgenic AD animal models. To achieve this aim, key objectives included building a curated dataset and making this dataset publicly available via an interactive web application. The protocol for a “living” SR (later

conceptualised as AD-SOLES) is available on the Open Science Framework (see Appendix C2.1).

2.3 Methods

2.3.1 Systematic literature search

Typically, search strategies for SRs are narrow and designed to answer a specific research question. I took a different approach and sought to first identify all primary research evidence related to Alzheimer's disease. To ensure that the search had the highest chance of encompassing all applicable studies, and to establish a database of studies containing AD research for use in future projects, I did not limit the search to animal studies. I searched PubMed, Embase and Web of Science Core Collection databases on 23rd January 2018 with broad Alzheimer's search terms (Table 2.2) and imported retrieved citations into Endnote X8.

Database	Search Terms
Web of Science (Core Collection)	TOPIC:(alzheimer)
PubMed	(Alzheimer's Disease[MeSH Terms] OR "alzheimers disease"[All Fields] OR alzheimer*[All Fields])
EMBASE (via Ovid)	alzheimer disease/ or alzheimer*.mp.

Table 2.2: Search strategy to identify relevant AD research

2.3.2 Duplicate removal

Initial deduplication (to remove duplicate citations obtained from multiple sources) was performed using Endnote X8, following recommended protocols (Bramer, Wichor M., Giustini, Dean, de Jonge, Gerdien B., Holland, Leslie, & Bekhuis, Tanja,

2016) followed by further manual sorting and removal of duplicates. Further duplicate removal was performed by ASySD, a novel tool I developed for removing duplicates in preclinical SRs (Chapter 4).

2.3.3 Crowdsourced reviewer training and engagement

I recruited a crowd of reviewers from within the EQIPD consortium and invited external collaborators to contribute. Fifty-five training papers were selected from the citations identified in the search. Two independent reviewers within our group (Myself and Emily Sena) annotated each paper and noted whether it involved primary research and then categorised it by study type, subjects, and, if relevant, further details relating to animal modelling. The full categorisation structure is shown in (Figure 2.1). We also determined whether each paper should be included or excluded based on the inclusion criteria shown in Table 2.3. We included studies which had primary research in a transgenic AD model. Conflicts were discussed to reach an agreement on "gold standard" annotations for each training publication. Reviewers signed up to the LearnToSyRF platform (http://learn.syrf.org.uk/), where they were asked to annotate papers (selected randomly) according to the annotation guide (Appendix C2.2). To complete the training, reviewers had to correctly screen and annotate 10 consecutive papers (scoring 80% or above). I delivered a webinar to introduce the screening process to EQIPD consortia members and encouraged users to sign up and begin training. To maintain

engagement and boost morale, frequent updates and progress reports were circulated amongst the reviewers (Figure 2.2).

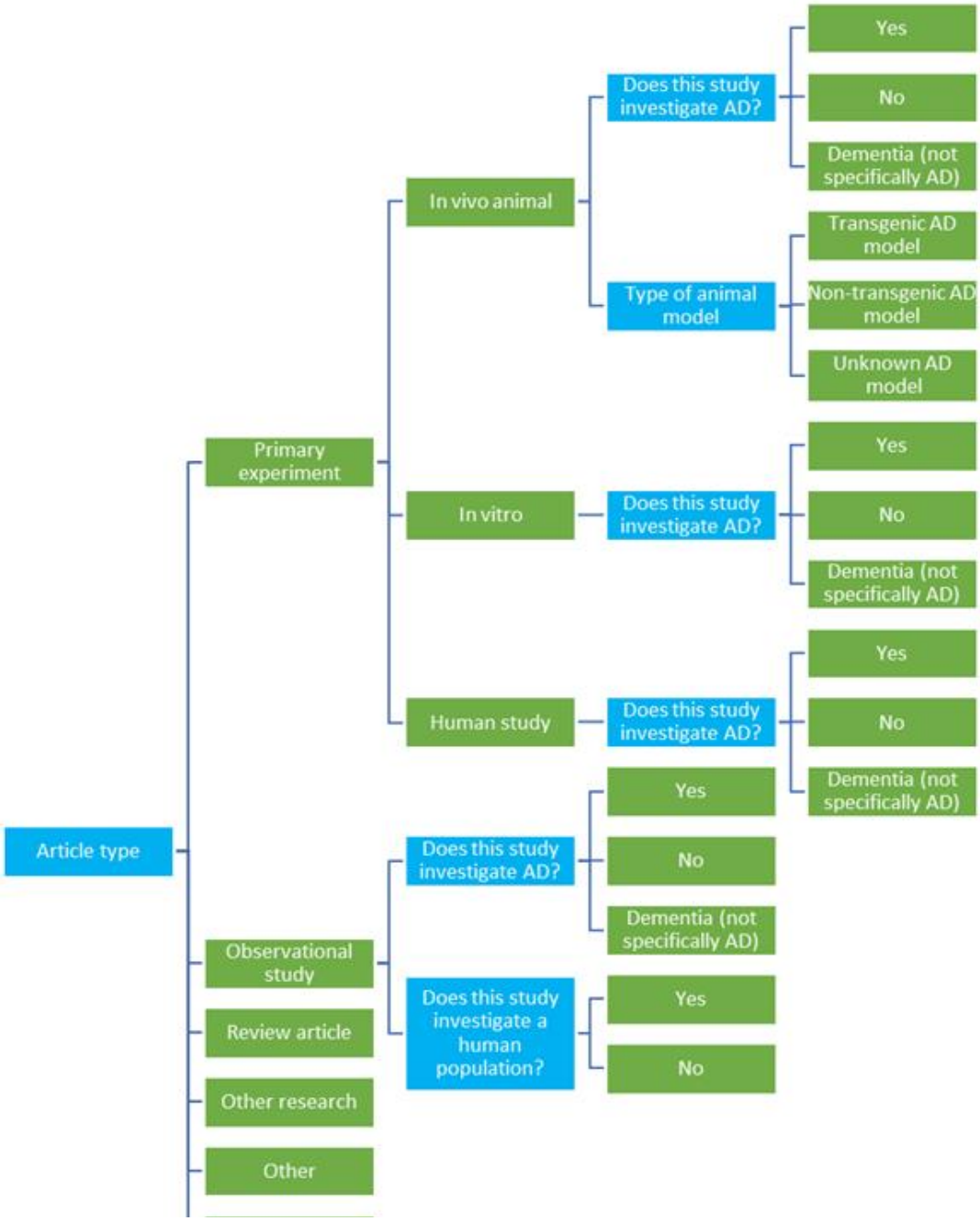


Figure 2.1: Categorisations used to annotate AD research publications

	Inclusion criteria	Exclusion criteria
Study design criteria	All publications including primary data from an AD animal model with an appropriate control group (wild-type littermate or appropriate control strain).	Publications without any primary data (e.g. reviews), publications without any primary data from animal AD models, publications with primary data in animal AD models without an appropriate control group.
Animal inclusion criteria	All publications with primary data describing an AD animal model of any species, age, or sex, including animals with comorbidity.	Publications with primary *in vitro,* or *in vivo* human data only.
Intervention inclusion criteria	All publications which model AD pathology by introducing one or more transgenes into their genetic sequence.	Studies which model AD pathology using other means (e.g. pharmacological interventions or environmental exposures) Toxicology studies and publications which have primary animal data only for wild type animals or other disease models.

Table 2.3: Inclusion and exclusion criteria for screening

WP2 SYSTEMATIC REVIEW PROGRESS REPORT: WEEK 10

Training Passes

Since we began, 17 contributors have signed up and 10 have passed the training at http://learn.syrf.org.uk. If you know anyone else who would be interested in getting involved, please get in touch!

Screening Leaderboard

Rank	Contributor	Papers Screened
1	Ann Marie	1120
2	Vincent	1006
3	Kaitlyn	907
4	Bettina	700
5	Isabel	340
6	Heidrun	305
7	Kim	166
8	Juan	102
9	Aneesha	100
10	Gillian	53
11	Malcolm	52
12	Emily	9

Screening Progress

We have made great progress over the past 10 weeks!

1718 papers have been screened by two independent reviewers, while **1451** papers have been screened by one reviewer. This is a grand total of **4887** screening instances, which is over **half** of our goal!

Thanks to everyone for their continuing progress and let's keep up the pace as we complete the final portion of papers!

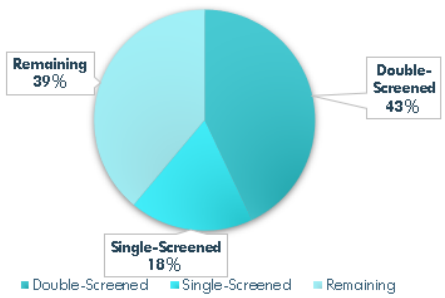


Figure 2.2: Example progress update sent to crowd of reviewers

2.3.4 Screening and annotation of subset

Using the sample_n function in R, a random sample of 4000 unique citations were selected and stratified by year of publication. These citations were then uploaded to a SyRF project.

Reviewers that had passed the training were then asked to categorise papers using prespecified criteria (see Figure 2.1). The SyRF platform allows users to annotate papers using a tree-like structure, with questions relevant to one categorisation (e.g. primary animal experiment) only appearing when this category is selected.

In collaboration with the University of Strathclyde, this dataset was also employed for an MSc project using categorised clinical AD research. After annotation by the crowd was complete, I coded each primary experiment conducted in humans as an randomised controlled trial (RCT) or non-RCT. Farah Francis acted as a second independent reviewer. All disagreements were discussed and rectified.

2.3.5 Reconciliation

Where two reviewers disagreed on whether a paper should be included, disagreements were reconciled by a third independent reviewer (Alexandra Bannach-Brown). Where there were disagreements on specific annotations, e.g. whether the record described an *in vivo* animal study or *in vitro* study, annotations were also reconciled by a third independent reviewer (myself, or Farah Francis where I was one of the original reviewers).

2.3.6 Machine-assisted screening

I used the annotated sample set of citations to train, validate, and test a ML classifier developed by, and hosted at, the EPPI Centre (University College London). This approach divides the sample set into a training set, validation set, and test set.

The machine selects relevant "features" or properties from the Tile and Abstract of each record in the dataset. These features are generated using a "bag-of-words" model, which transforms text into (i) word frequency counts, and (ii) additional numeric representations of word relevance and significance, so as to differentiate between records. These representations are used to train a classifier to distinguish between records. Firstly, the classifier is applied to the validation set, and the machine classification (Include/Exclude) is then compared to the "gold-standard" human classification (Include/Exclude). At this stage, the results are used to fine-tune parameters in the classifier and make additional improvements before the classifier is tested on an unseen test set of records. The comparison of human and machine classifications on the test set is used to evaluate the performance of the ML classifier.

To ensure maximum attainment of relevant studies, I predetermined that performance on the test set would have to reach a sensitivity threshold of at least 0.95, where:

$$Sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + \ number\ of\ false\ negatives}$$

True positives here indicate included studies, while false negatives indicate wrongly excluded studies. Where the sensitivity standard is reached, 95% of papers that should be included (according to Human screening decisions) are correctly classified by the algorithm as "Included". This is comparable to expert human performance, with the inter-screener agreement rate in SRs at CAMARADES estimated at between 95% and 99%.

As the sensitivity of a classifier increases, there is a risk of over-inclusion that must be balanced by a degree of specificity, defined as:

$$Specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$$

A specificity of 95% would, therefore, mean that 95% of papers that should be marked as "Excluded" have been excluded by the machine. Positive predictive value, or precision was also assessed. Precision is defined as:

$$Precision = \frac{number\ of\ positives}{number\ of\ true\ positives + number\ of\ false\ positives}$$

During the screening stage, it was clear that the inclusion rate was low. To provide more "Included" instances to train the ML classifier, I enriched the dataset with 427 relevant studies from a previous SR of transgenic AD models (Egan and Macleod, 2016). Some papers did not have an abstract, and reviewers could not reliably screen these studies for inclusion or annotate them based on the title alone. Upon removing these papers, machine performance improved substantially, and I therefore decided to remove papers for which we could not find an abstract from this stage of the review.

2.3.7 Error analysis

Error analysis methods can flag human errors by identifying records where the human and machine decisions are mismatched. This allows for errors to be corrected, and can result in performance improvements in the screening algorithm (Bannach-Brown et al., 2019). To perform error correction, 400 records where the machine disagreed most with the human assessment (e.g. the machine assigned a high value to a record to indicate it should be included but a human marked it as excluded) were selected. I checked that the human annotation for each of these

records was correct and corrected where appropriate, before training the machine again and applying the algorithm to the entire AD dataset.

2.3.8 PDF retrieval and updating via Endnote

I used the Endnote X8 "find full text" feature to retrieve full-text PDFs using ezproxy links to access the University of Edinburgh subscriptions. Using the Endnote X8 "update references" function, I attempted to retrieve additional information for every record in the dataset, including abstract information if available.

2.3.9 Machine-assisted annotation and risk of bias assessment

To enable users of AD-SOLES to quickly identify relevant research of interest, I extracted the animal model(s), intervention(s), and outcome measure(s) or "MIO" elements within publications using a text-mining approach (see Chapter 3 for details and validation).

I created regular expression (regex) dictionaries for transgenic models based on a well-maintained list of animal models of AD on the Alzforum website (https://www.alzforum.org/research-models/alzheimers-disease). Synonyms are given for most models, and I went through each model individually to add alternative names where appropriate. Similarly, for therapeutic interventions, I based my list on the therapeutics database within the Alzforum platform (https://www.alzforum.org/therapeutics) and added alternative names for compounds when provided. To develop outcome measure specific regexes, I worked from a list of commonly used outcome measures specified in a previous SR (Egan et al., 2016), with a particular focus on the outcome measures used in my AD SR projects (Chapter 5, Chapter 6). A link to the regex dictionaries is available in Appendix C2.5

I built upon existing regex approaches (Bahor et al., 2017) to assess the reporting of measures to reduce the risk of bias (RoB), e.g. randomisation of animals to experimental groups, blinded outcome assessment, performing a sample size calculation, and conflict of interest statements (see Chapter 3). I also extracted the country where a research study was conducted using data from the "author address" field.

To apply regex techniques, using the AutoAnnotator R package (Liao, 2017), PDFs are first converted to text using pdftotext (a freely available, open source tool available at: https://www.xpdfreader.com/pdftotext-man.html).

2.3.10 Building an R Shiny web application

R Shiny (https://shiny.rstudio.com/) is an R package which allows users to create webpages and dashboards written in R code, so users can directly interact with datasets. Once the dataset of transgenic AD studies had been collected and annotated for model, treatment, outcome, county, and RoB reporting, I developed a web application to display the dataset visually and allow users to interrogate it. I created visualisations to summarise the country of corresponding author, the estimated proportion of research conducted across different models, testing different treatments, and measuring different outcomes. I added RoB assessments throughout, so that users could look at the quality of existing research across different models/interventions and added date filters so that users could look at changes in research methods and reporting quality over time. All citation information was included within the application, so that users could apply filters (e.g. selecting a specific transgenic model) and download curated citation lists of relevant research for further evaluation. A link to the R code underlying the shiny web application is available in Appendix C2.4.

2.4 Results

2.4.1 Systematic search

In January 2018, I identified 131,992, 181,673, and 127,939 records from PubMed, EMBASE, and Web of Science respectively, totalling 441,604 records. Duplicate references were removed following established deduplication steps in Endnote X8 (Bramer, Wichor M. et al., 2016) followed by further manual sorting and removal of duplicates by hand. Of those 441,604 records, 265,258 remained in the dataset after deduplication. A subgroup was then randomly selected for screening and annotation by the crowd. Due to the vast number of records and limitations of Endnote, many duplicates were missed at this stage. Following further deduplication using ASySD, the AD dataset contained 230,203 unique records. Finally, after the removal of references without an abstract, 184,333 records remained.

When duplicate references and references without abstracts were removed from the sample set of 4,000 papers selected for screening and annotation by the crowd, 3,264 records remained.

2.4.2 Crowdsourcing and engagement

A total of 21 (including myself) reviewers signed up to train on the LearnToSyRF platform. Of these, 14 were members of the EQIPD consortium. Sixteen reviewers passed the training and contributed to the screening and annotation of AD research.

2.4.3 Screening for inclusion

Reconciliation was required on 189 papers where reviewers had disagreed on whether it should be included or excluded. A total of 347/4000 papers were

included. When duplicates and records without abstracts were removed, the final sample set comprised 260 included and 3004 excluded citations.

2.4.4 Categorisation of studies

Reviewers fully agreed on their categorisations for 1,792/3,264 papers in the final sample set. During the reconciliation process, several issues caused a high level of disagreement on certain annotations, specifically:

1. Article type as “Review” article versus “Other”: one reviewer may classify non-primary research as “Review”, although there is no evidence from the abstract – the paper could be an opinion, commentary, or letter.

2. Article type as “Primary research (*in vitro* experiment)” versus “Other research”: most reviewers lacked specialist knowledge of *in vitro* AD models or paradigms and there was disagreement about whether, for example, Aß aggregation outside of any cell would be classed as a chemistry-like study (“Other research”) or an *in vitro* study relevant to the pathophysiology of AD.

3. Whether a study investigated AD: many studies could be relevant to AD in some way (e.g. the mental health of AD caregivers, the chemical structure of Aß plaques, or the attitudes to AD in society), but do not investigate disease pathology in human patients with suspected AD or laboratory AD models.

Taking these considerations into account, I combined “Other” and “Review” into one larger category of “Secondary research” and did not use “investigates AD” as a category.

In the final reconciled dataset, the largest proportion of publications in the subset were observational studies. A summary of publications in each category in the final reconciled dataset are shown in Table 2.4. Many publications fell into more than

one category, as shown in Figure 2.3. Overall, the largest proportion of publications in the subset were observational studies, followed by secondary research. For publications categorised as primary research, most described *in vivo* animal studies in transgenic AD models. For a small number of publications (n=18), it was unclear to reviewers whether the animal model used was a transgenic model or if the AD pathology was induced by another method.

Article category	Subcategory (if applicable)	Number of Papers
Observational study	**Total**	**1,253**
	In human subjects	1,215
	Other research types	38
Secondary research	**Total**	**879**
Other research	**Total**	**434**
Primary experiment	**Total**	**846**
	In vivo animal	441
	I*n vivo* animal (AD model)	311
	In vivo animal (Transgenic AD model)	211
	I*n vivo* animal (Non-transgenic AD model)	86
	In vivo animal (Unknown AD model)	18
	I*n vitro/ ex vivo*	383
	In vivo human	105
	In vivo human RCT	35
	In vivo (can't tell if human or animal)	3
	Unknown	2
Unknown	**Total**	**36**

Table 2.4: Number of publications categorised into each research type

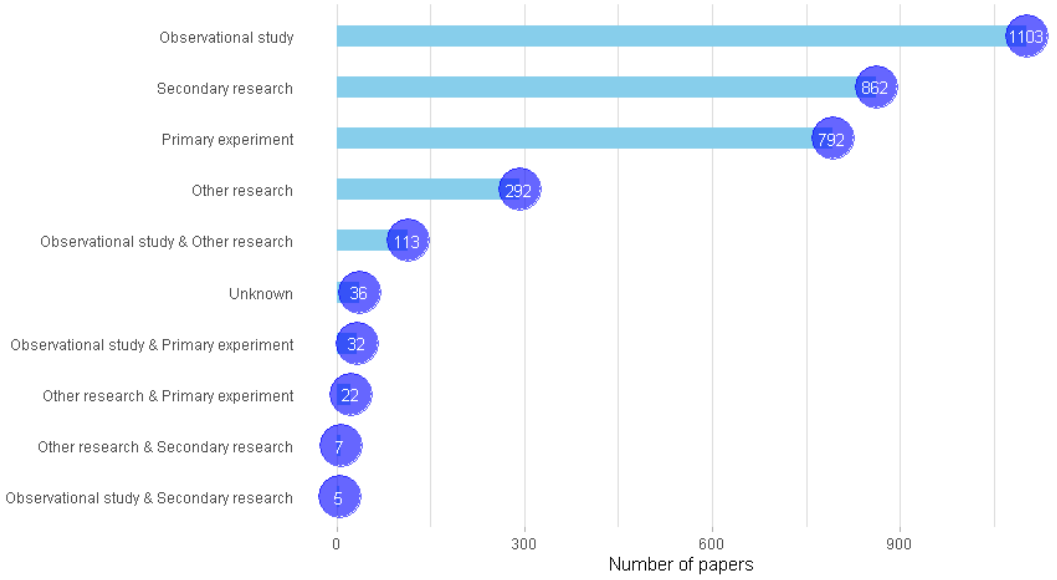


Figure 2.3: Lollipop plot of publication research types (including combined categorisations)

2.4.5 Machine-assisted screening algorithm

I obtained 427 relevant records from a previous SR of transgenic AD models (Egan and Macleod, 2016). Of these, 422 were still available online, and of these only 389 had abstracts.

The final dataset used to train, validate, and test the ML algorithm therefore comprised of 3,653 records (3,264 from the crowdsourced annotation + 389 from the previous AD SR). Of these, 649 records (260 from crowdsourced annotation + 389 from previous AD review) were labelled as “Included” and 3004 were labelled as “Excluded”.

The algorithm was trained and validated on 2999 papers in the screened dataset before being applied to the test set of 654 papers to evaluate performance (Table 2.5). Prior to error correction, the machine performed at a sensitivity (proportion of records correctly included) of 0.962 and specificity (proportion of records correctly

excluded) of 0.854. Of the 400 human-machine disagreements identified for error analysis, seven true human errors were corrected. Post-error correction, sensitivity increased to 0.965 and specificity increased to 0.878. The precision of the final classifier was 0.429. The algorithm assigned numerical values in the likelihood of inclusion ranging from 0 (confident "Included") and 1 (confident "Excluded") then selected the best performing threshold for partitioning the data into "Included" and "Excluded" records. As shown in Figure 2.4, the best threshold to distinguish the records was a score over 0.14.

	Human: INCLUDED	Human: EXCLUDED
ML: INCLUDED	55	73
ML: EXCUDED	2	524

Table 2.5: Confusion matrix from test set (N = 649 records)

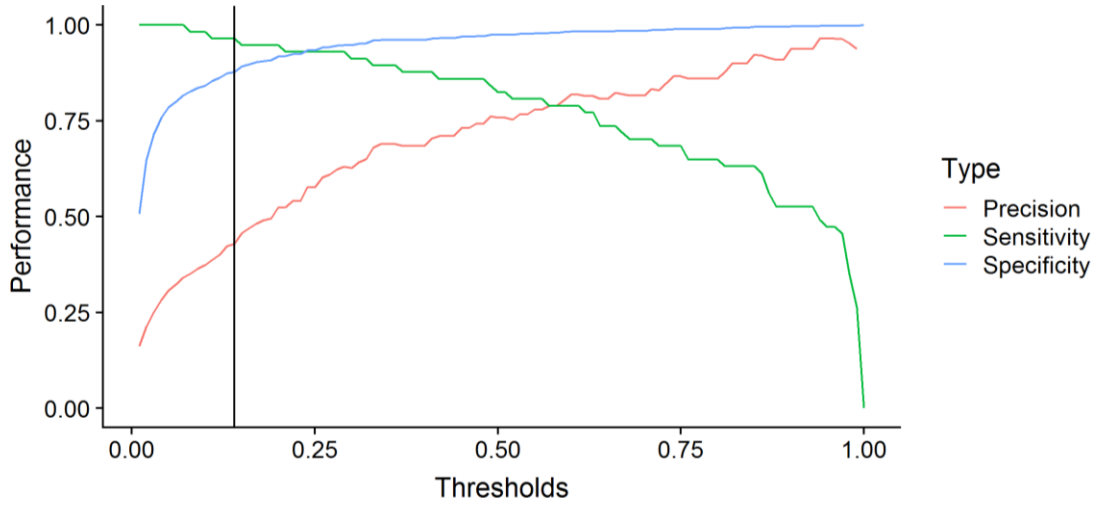


Figure 2.4: Selection of best threshold for ML classification of included studies

The performance of the algorithm was above the "gold-standard" 95% sensitivity threshold and was therefore applied to the remainder of the dataset (N=181,069 without the sample set). The algorithm subsequently included 26,672/184,333 total records. See Figure 2.5 for an adapted PRISMA flowchart detailing the number of

records at each stage (Moher, Liberati, Tetzlaff, & Altman, 2009). A link to the list of included publications is shown in Appendix C2.3.

2.4.6 PDF retrieval and updating via Endnote

Using Endnote X8, I obtained 84.1% (22,432/26,672) of the fulltext PDFs for the included studies. Using the "Update References" feature improved PDF and abstract retrieval. I discovered that 7 records had been wrongly updated – with changes to title/abstract information. This caused considerable disruption as some studies had to be re-screened with the correct abstract information, and reference information had to be matched and changed back to the original. It is important to be aware of this Endnote failure in future reviews. It can be largely rectified by restricting Endnote from changing the whole reference, and instead selecting the option to "update only missing information".

2.4.7 PDF conversion to text for machine-assisted annotation

The pdftotext program successfully converted 99% (22,375/ 22,432) of publications to text files. Therefore, 22,375 records were assessed using regex tools to extract MIO information and RoB reporting.

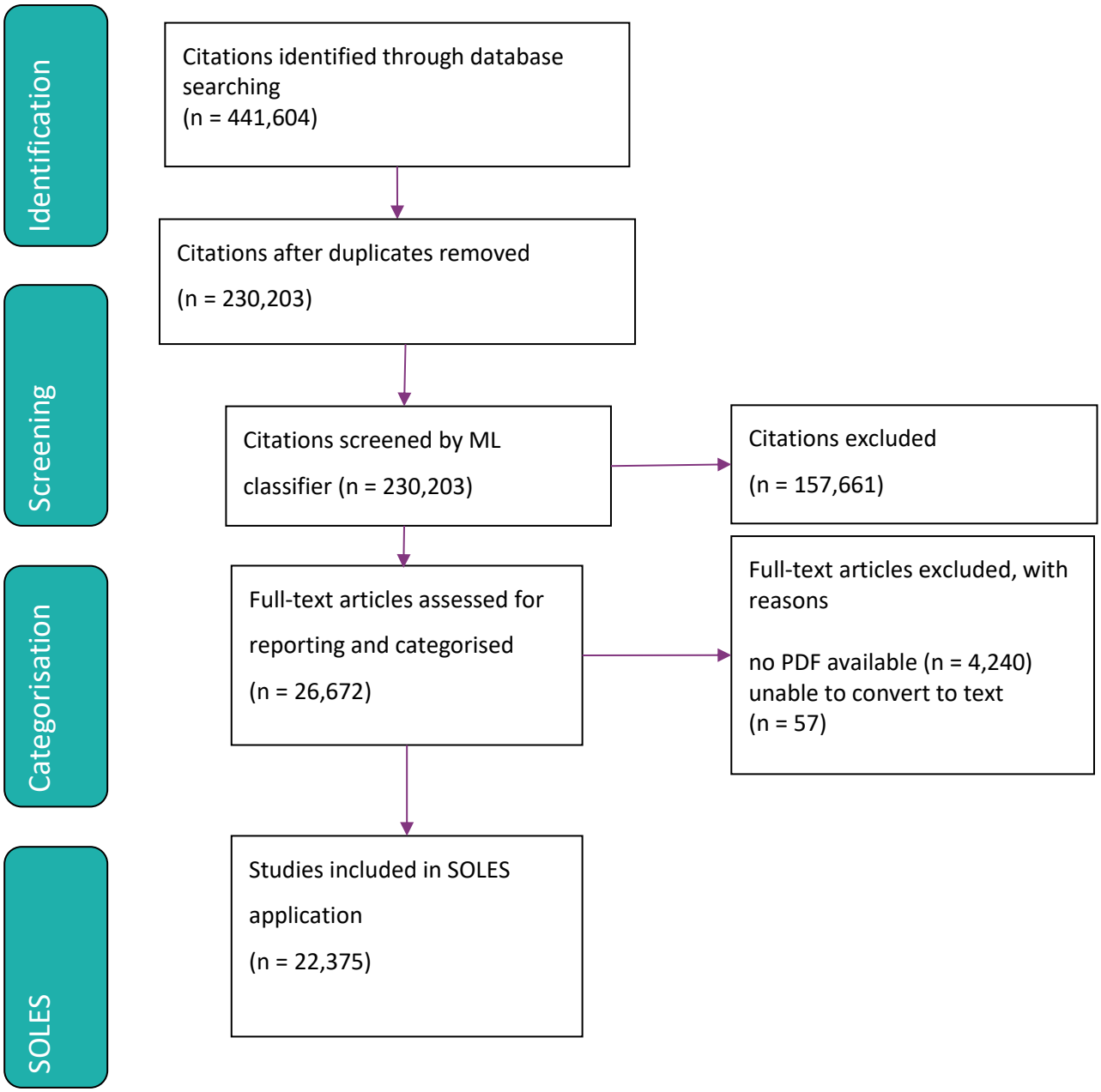


Figure 2.5: PRISMA flow diagram for AD-SOLES pilot

2.4.8 Visualisation and Web Application

The resulting data were visualised in an interactive online application developed using R Shiny (see link in Appendix C2.6). The first page of the application shows the number of publications in the dataset over time, showing the number of new papers published each year (bar plot) and a cumulative count. As seen in the screenshot (Figure 2.6), the number of studies rose substantially in 2009 and has continued at a rate of approximately 1,800 new papers a year. A choropleth map shows the number of publications by country, and a bar plot shows the overall reporting of key RoB items (blinded outcome assessment, randomisation to experimental groups, conflict of interest statements and sample size calculations). Users can hover over plots to see exact numerical values for any data point. Overall, RoB reporting across the dataset is poor, with only 14.6% of papers reporting blinded outcome assessment, 27.6% reporting conflicts of interest, 18.54% reporting randomisation, and 4.0% reporting a sample size calculation. It was not possible to attain country information for every publication due to missing author information, but of those I could extract, researchers in China had published the most research in transgenic AD models, followed by Germany.

The Modelling, Interventions, and Outcomes pages are similar (Figures 2.7-2.9). Users can select a year range of interest and a sunburst chart with the number of publications in each broad category (e.g. intervention target, gene mutation(s), and outcome category), subcategory (transgenic model, drug, and outcome) is displayed. On the right-hand side, the RoB reporting for each broad category is shown. When a broad category is clicked, the sunburst chart expands, and the reporting quality bar plot shows reporting quality for each subcategory (Figure 2.10). Similarly, a specific MIO element can be selected (e.g. the recently improved anti-AD drug Aducanumab in Figure 2.10) and the reporting quality of potentially relevant studies is displayed (Figure 2.11).

On the Download Citations page, users can filter by year, model, intervention, and outcome to view and download citations for relevant references (Figure 2.12).

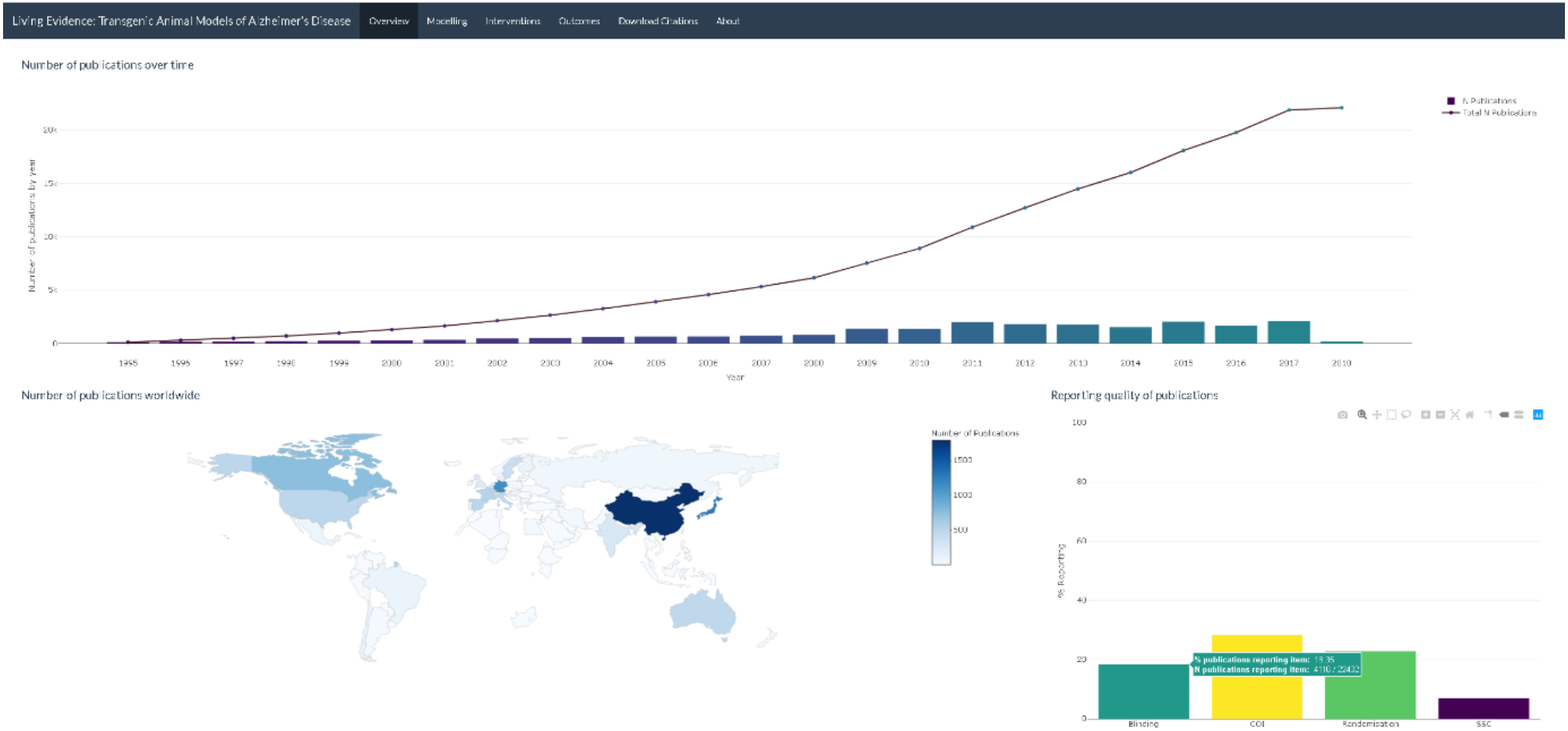


Figure 2.6: Screenshot from AD Shiny application overview page

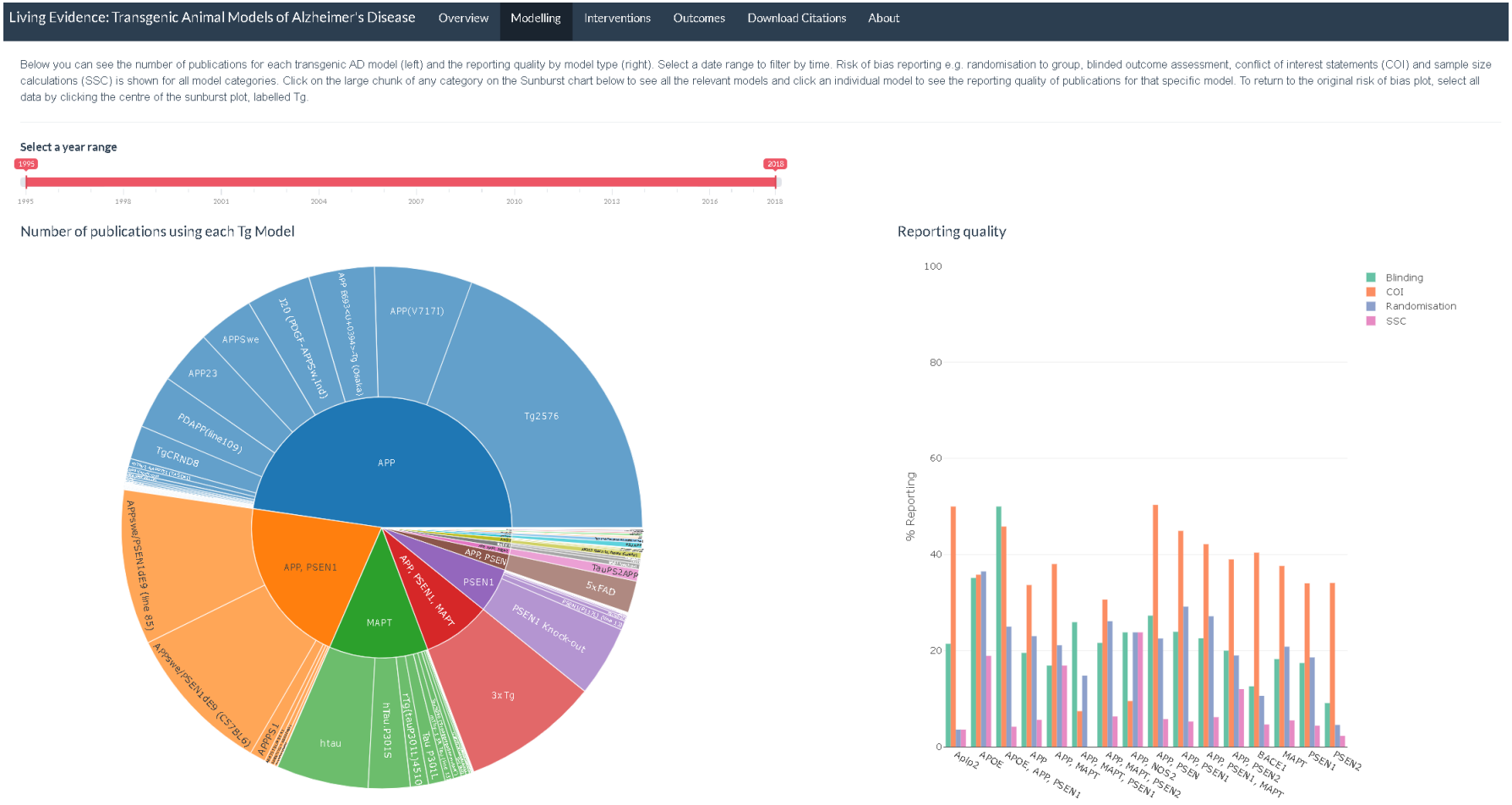


Figure 2.7: Screenshot from AD Shiny application modelling page

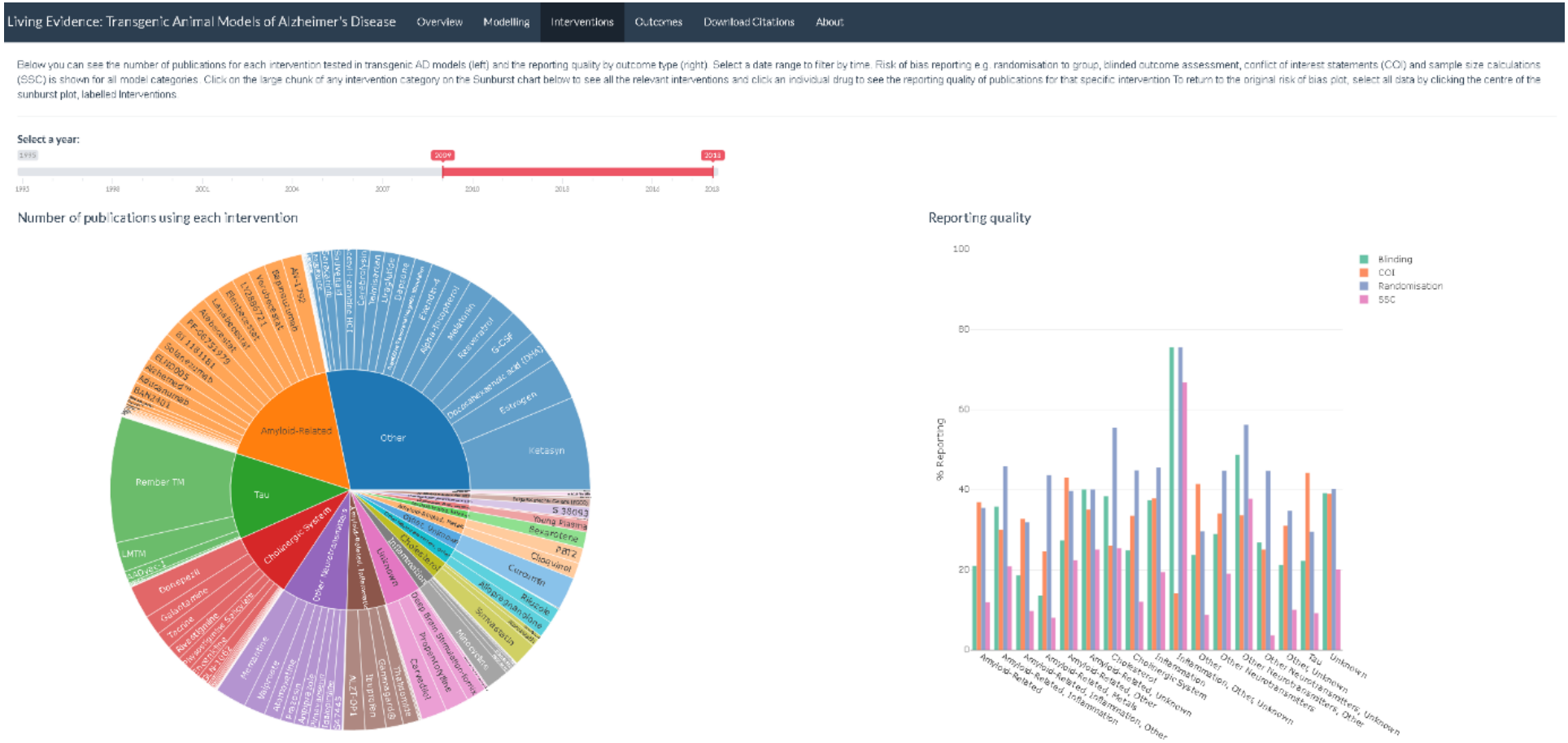


Figure 2.8: Screenshot from AD Shiny application interventions page

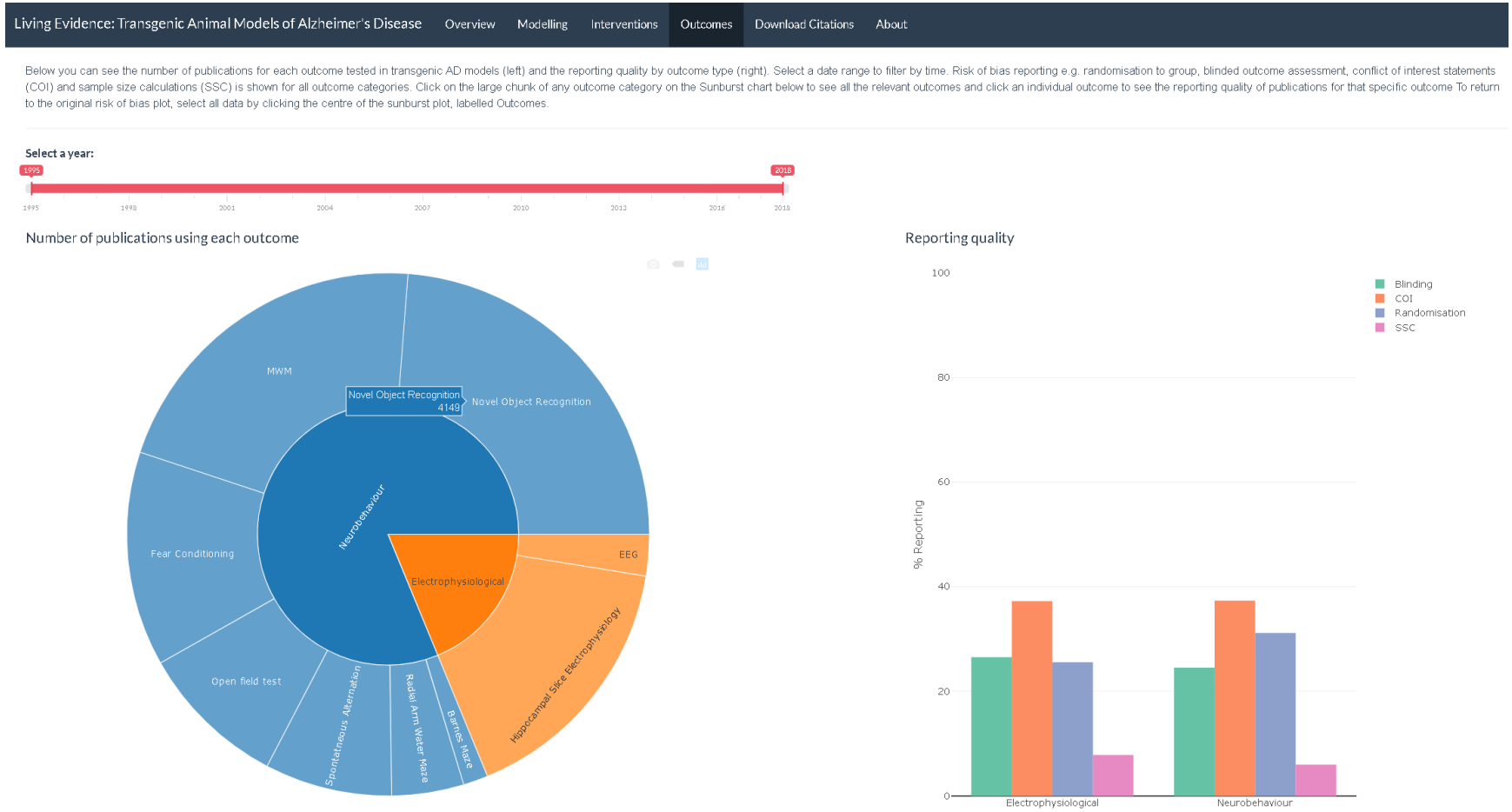


Figure 2.9: Screenshot from AD Shiny application outcomes page

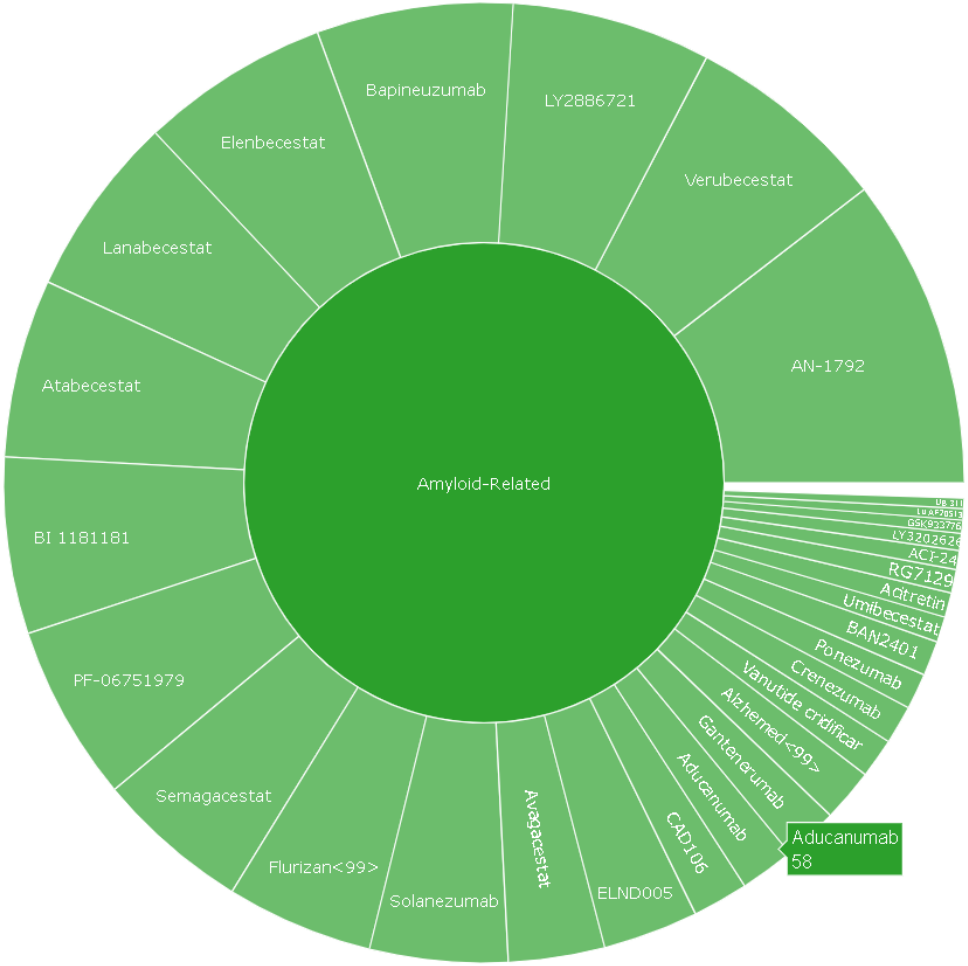


Figure 2.10: Screenshot from AD Shiny application - expansion of intervention sunburst plot

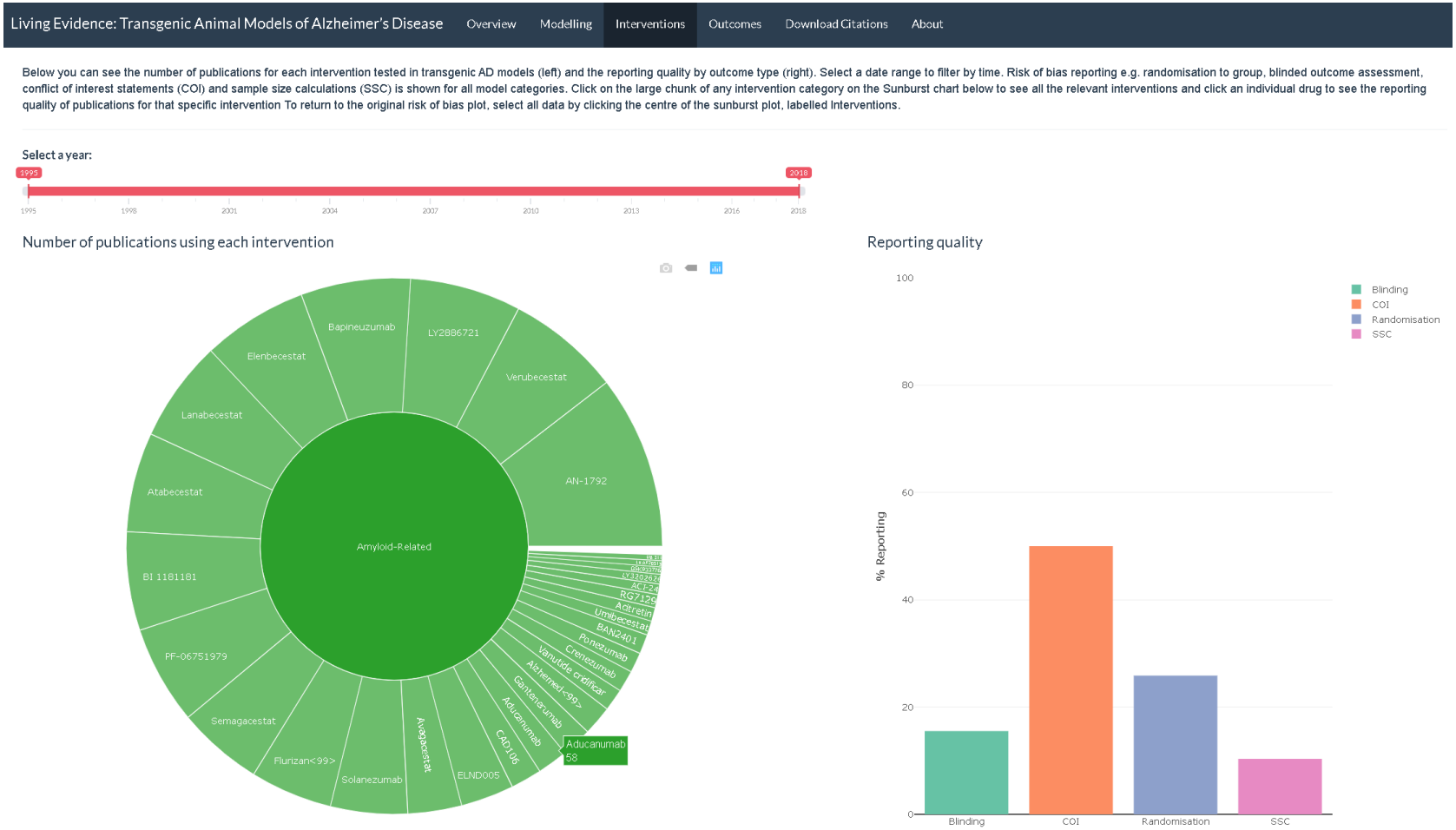


Figure 2.11: Screenshot from AD Shiny application: risk of bias reporting for a specific intervention

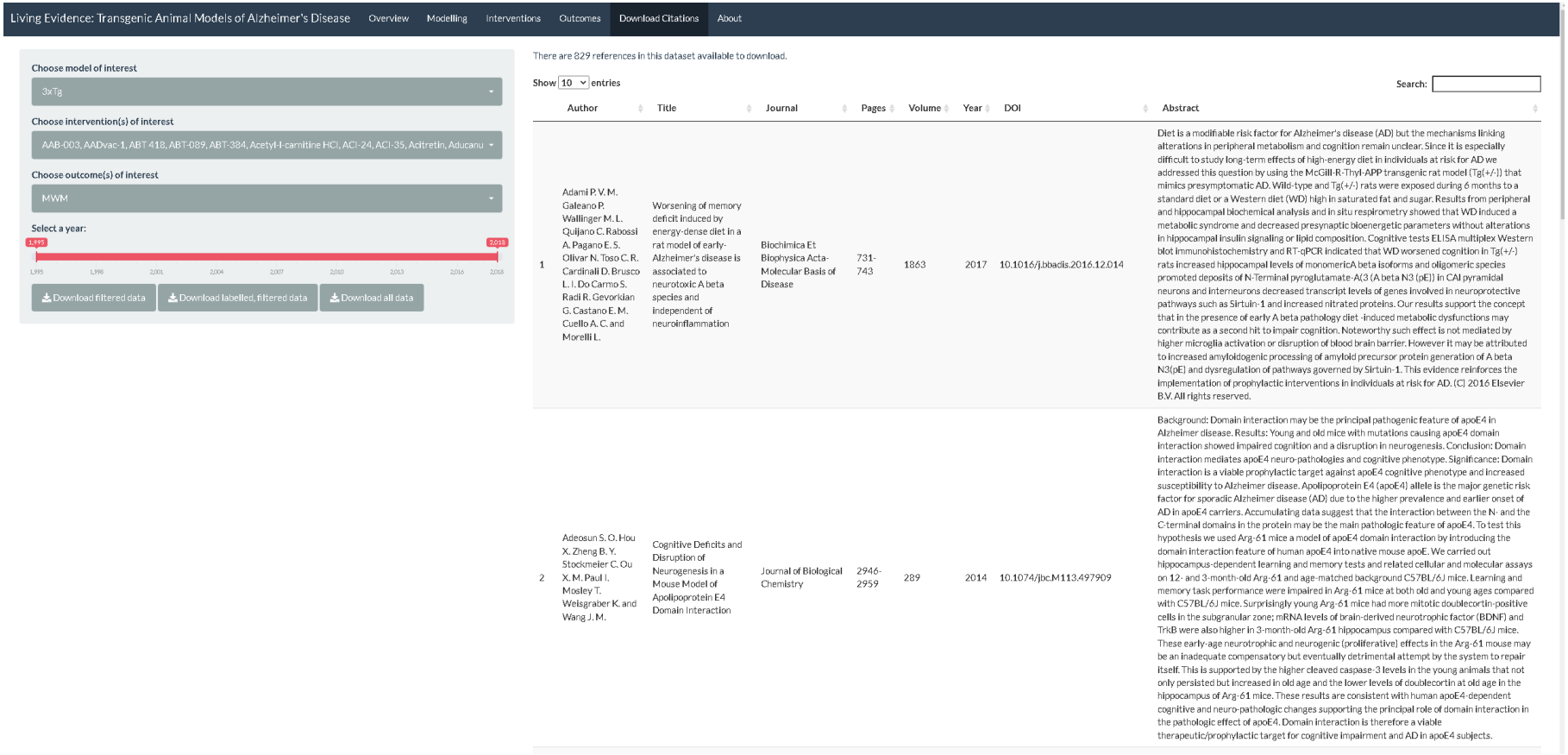


Figure 2.12: Screenshot of AD shiny application downloads page

2.5 Discussion

2.1.2 AD-SOLES

This project has delivered a curated synthesis of the evidence from preclinical AD research. This framework has served as a starting point to facilitate a series of SR and meta-analysis projects investigating the open field paradigm and synaptic plasticity and memory deficits in transgenic AD models (See Chapters 5 and 6). Through the shiny web application, AD-SOLES provides all research groups with an opportunity to perform additional reviews of the literature at an accelerated pace.

This was a pilot project to demonstrate the feasibility of AD-SOLES, and I intend to develop this, in collaboration with the AD research community. By providing an overview of what is currently known, and by identifying which variables are of key importance to experimental rigor, this approach could inform, guide, and improve future preclinical trials, bridging the gap to translation.

The SOLES approach addresses several of the key barriers laid out in Chapter 1 which impede the impact of preclinical SRs. Through automation and crowdsourcing approaches, I have summarised evidence from thousands of publications using transgenic AD models, which would not have been possible using conventional approaches. Further, by using a broad, all-encompassing search strategy, I hope to have collected the vast majority of available literature. Within the subset of citation title and abstracts annotated by human reviewers, the type of animal AD model used (transgenic vs non-transgenic) was unclear in a small number of publications. These publications were still classed as included, as they may have contained relevant research in transgenic AD models. It is therefore likely that performing a systematic search for transgenic models alone would have missed a small but significant proportion of the literature.

AD-SOLES has the potential to benefit primary researchers, who can use the platform to inform future research, and to significantly accelerate their own SR and

meta-analysis projects. Stakeholders (e.g. funders and institutions) could use AD-SOLES to gain a clearer insight into the research landscape, benchmark quality, and also to prioritise areas for improvement or further research. Clinical trialists could use the platform to evaluate available evidence, collect the relevant citations, and determine whether progression to a clinical trial is appropriate in a certain drug. Developers of evidence synthesis technologies could use the underlying dataset to validate new tools or apply the framework to other research areas.

2.1.3 Limitations

In its current state, AD-SOLES is not "living" as it does not hold studies published beyond 2018. Ideally, I would implement a fully automated workflow that would retrieve new records every day, categorise them, assess reporting, and visualise the results in a versioned shiny app. A major roadblock in this pipeline is automated search retrieval. PubMed has an accessible API that allows for this. Unfortunately, implementing automated citation retrieval from Embase, Web of Science has not been so straightforward and requires additional fees and subscriptions. To advance to a "living" evidence summary, automation of the database searches could be achieved by either limiting citation retrieval to easily accessible databases (including PubMed and preprint databases like Biorxiv and Medrxiv), gaining access to these APIs, or potentially utilising RSS feeds. Collaboration with data engineers and software developers could help enable this functionality.

Another key limitation in the AD-SOLES workflow is the omission of potentially relevant records at multiple points. Twenty percent of the final unique dataset (45,870 / 230,203 records) lacked abstracts and were removed as the machine classifier could not distinguish papers from a title alone. It was not within the scope of this project to investigate or classify records lacking abstracts, and it is unclear if records in this category share similarities or belong to a certain category of research. Endnote X8, using university subscriptions, retrieved PDFs for 84%

(22,432/26,672 records) of included studies. Incomplete PDF retrieval impeded the ability to automate the classification of publications, as our current text-mining tools could not assess quality or extract model, intervention, and outcome measures without having access to the full-text. Within the timeframe of the project, it was not possible to manually search for missing PDFs, so it remains relatively unknown whether this was a result of an issue affecting Endnote X8's retrieval or rather our university's subscription not covering those references. More generally, the culture shift towards open access publishing practices will continue to improve access to full-texts. The machine classifier was optimised for sensitivity over specificity, and while I am confident that it included as many relevant records as possible, the process was over-inclusive. The classification of MIO elements with regex dictionaties was also designed to be highly sensitive and may require further adjustment and validation (Chapter 3).

2.1.4 Future Directions

The SOLES methodology is flexible and can be applied to different research domains. Once a "living" workflow is fully realised, my next step for this project would be to categorise all *in vivo* primary studies using animal models of AD, with later plans to expand to all relevant *in vitro* and human AD research.

To build upon the database of human-annotated papers, and ultimately improve the performance of automated tools in future SOLES projects, I could adapt the training materials and approaches from this project to attract and train a larger crowd. Uptake could be improved if I formalised this approach and developed an attractive and user-friendly interface (Nama et al., 2019), such as Cochrane Crowd for clinical SRs (Noel-Storr et al., 2017). Given the relevance to dementia, I could also promote involvement in this project at relevant conferences and networking events. With input from leading research groups, I could refine AD-SOLES to ensure it captures the most clinically relevant information.

If I engage the community and provide training in SR methodologies, this may also lead to more preclinical SRs being conducted. All reviews derived from AD-SOLES could be linked on the web application to facilitate visibility and reduce the risk of unnecessary repetition. If researchers who perform preclinical SRs later share the data they have collected, this could also facilitate the creation and validation of more automated tools to enhance the speed and capabilities of meta-research approaches.

Further, involving laboratory researchers more directly and attaining subject-specific guidance could aid in the development of new tools to extract the most relevant and useful information from publications. Through continuous monitoring of key aspects of rigour and reporting quality over time, guided by findings from AD SRs, the ARRIVE guidelines, and disease-specific expert guidance, AD-SOLES could provide insights into areas where improvement is required. This could lead to the development of specialised "living" guidelines (Akl, Meerpohl, Elliott, Kahale, & Schünemann, 2017), research improvement targets, and initiatives to maximise the validity, transparency, and reproducibility of *in vivo* experiments, while minimising research waste.

2.6 Conclusion

Making sense of immense fields of preclinical research is an ever increasing challenge. Without effective processes for collecting, critically appraising, and synthesising the evidence, incorrect decisions may be made, repetitive research studies performed, and ultimately time, research funding, and laboratory animals wasted. Using a combination of crowdsourcing and automation approaches, this chapter has outlined a novel approach to creating the foundations of a systematic online living evidence summary of preclinical AD research.

CHAPTER 3: DEVELOPING TEXT-MINING TOOLS TO SUPPORT FASTER EVIDENCE SELECTION

3.1 Chapter Introduction

In this chapter, I validate the use of automated text-mining approaches to extract methodological details from publications without the need for a human reviewer. These methods were used extensively to create the annotated dataset underlying the AD-SOLES pilot project, described in Chapter 2.

3.2 Background

3.2.1 Text-mining approaches in SRs

Text-mining is the process of deriving high-quality information from text using automated or semi-automated methods (Hearst, 1999). In the context of SRs, text-mining approaches are still relatively novel, but in recent years there have been significant advancements in their application to different stages of the review process (O'Mara-Eves, Thomas, McNaught, Miwa, & Ananiadou, 2015). For example, text-mining approaches have been proposed to improve the sensitivity of systematic search strategies by extracting relevant keywords from within publications (Ananiadou, Rea, Okazaki, Procter, & Thomas, 2009). As discussed in Chapter 2, text-mining approaches utilising trained machine learning classifiers can be used to attach "include" and "exclude" decisions to studies based on text within the title/abstract, significantly reducing the time and resources required during screening (Bannach-Brown et al., 2019; Cohen, 2011; Liao et al., 2018; Wallace, Trikalinos, Lau, Brodley, & Schmid, 2010). Complimentary approaches have been developed to prioritise representative publications for screening earlier than others, which offers potential time-saving benefits as machine classifiers can be trained on the most relevant publications first (Ananiadou et al., 2009). Prioritisation approaches may also allow researchers to get a sense of the field more quickly than

screening publications from the pool at random (Cohen, Ambert, & McDonagh, 2009).

Much of these developments have been aimed at clinical literature, with less emphasis on preclinical animal research. In addition, less attention has been paid to applying text-mining approaches to automate aspects of data extraction and quality assessment. In this chapter, I describe text-mining approaches which aim to:
(1) extract MIO (model(s), intervention(s), outcome(s)) information from publications to enhance evidence selection, and
(2) extract RoB reporting from publications.

3.2.2 PICO extraction to enhance evidence selection

To summarise and evaluate biomedical research in an unbiased way, a systematic search strategy should be sensitive enough to identify all potentially relevant research. However, a recent analysis of 195 prior SRs observed that, on average, more than 75% of papers identified in the search were later excluded (Borah et al., 2017). Given that it can take a skilled reviewer between 30 seconds and several minutes to assess a citation for inclusion during screening (Wallace et al., 2010), the retention of irrelevant citations is extremely costly, both in time and resources. To summarise evidence in shorter timeframes, rapid review methodology has been developed, placing a higher value on specificity. As such, rapid reviews are not fully inclusive and can omit relevant evidence (Watt et al., 2008). A sensitive approach to identifying as many studies as possible combined with an ever-growing body of literature poses an escalating logistical challenge.

SRs, in general, identify potentially relevant records by searching biomedical databases for key terms relating to the population(s), intervention(s), comparator(s), and outcome measure(s) of interest (hereinafter “PICO”). In the preclinical domain, where researchers have greater control over experimental

design, the comparator or control group of animals does not often, in our experience, form part of a systematic search strategy. Instead, at the screening or data extraction stage, a reviewer may decide whether an experiment has an appropriate control group. For this reason, I have conceptualised a modified PICO-like concept for preclinical SRs: MIO (model, intervention, outcome).

Most databases allow users to search certain bibliographic fields such as title, abstract, journal, and keywords. Few databases allow users to search through the full-text of publications. Therefore, the sensitivity of current approaches relies upon the presence of clear PICO/MIO terms in the publication title or abstract.

The balance of needing to identify all relevant research and the feasibility of completing a review in a timely manner is difficult to resolve using traditional approaches. Applying automated methodologies – such as text-mining - to the review process has the potential to reduce the human workload required, without substantial sacrifices in sensitivity. Several groups have developed tools to automatically extract PICO-like statements from within the title/abstract of publications (Chung, 2009; Kim, Martinez, Cavedon, & Yencken, 2011).

However, this approach may be less useful within the preclinical domain. Anecdotal experience within our group suggests that MIO information is often omitted from the abstracts of preclinical studies, possibly due to there being multiple animal models used, different treatments tested, and/or several outcomes measured within one publication. Extracting MIO elements from the full-text might prove a more useful approach.

Researchers have trained a machine classifier on clinical trial summaries, developing a tool to automatically extract PICO elements from the full-texts of clinical trial reports (Wallace, Kuiper, Sharma, Zhu, & Marshall, 2016). Recent work within the CAMARADES group has utilised regular expressions (regex) to extract MIO elements

from the full-text of preclinical studies (Bannach-Brown et al., 2021). Regexes are specialised character sequences that comprise a search pattern. Such patterns are case sensitive and can incorporate both standard syntax (text characters) and meta-characters which have special meaning. Figure 3.1 breaks down one example of a simple regex search pattern to identify the use of the Morris water maze behavioural test within a publication.

Using a regex approach, one can develop dictionaries for commonly used models, drugs, and outcome measures which can be run against the full-text publication. An R package, AutoAnnotation (Liao, 2017) was developed within our group to count the frequency of matches within the full-text to each dictionary term. However, there is still some uncertainty around the accuracy of such approaches, and what threshold should be set i.e., how many matches with "Morris water maze" indicates that the publication reports Morris water maze outcomes. Further, there is no clear subdivision between sections of the paper, so the regex may match something within the reference section of a publication that is irrelevant to the experiments it reports.

Regular expression	**Matching text**
(Morris)?	"**Morris**" 0 or 1 times
[Ww]ater	"**Water**" or "**water**"
[Wwater].[Mm]aze	"**Water maze**" **or** "**water-maze**" (. matches any character)
(Morris.)?[Ww]ater.[Mm]aze	"**Morris Water Maze**" or "**Water maze**" or "**water maze**" (matching with or without "Morris ")
(Morris.)?[Ww]ater.[Mm]aze|MWM	"**Morris Water Maze**" or "**Water maze**" or "**water maze**" or "**MWM**"

Figure 3.1: Example regex pattern to identify Morris water maze outcomes

In this work, I aimed to validate regex-based approaches to identify the transgenic model and outcome measures used in the preclinical AD literature. I also aimed to evaluate the utility of such approaches versus specialised systematic search strings

which rely on the clarity of title/abstracts. Classifying publications into different MIO categories supports the SOLES approach (Chapter 2) of synthesising research evidence.

3.2.3 Automated RoB assessment

As discussed in Chapter 1, findings from preclinical experiments may overstate the efficacy of interventions as they are at risk of systematic bias. Measuring the reporting of measures to reduce the RoB within a given research area is therefore important when evaluating the strength of available evidence. Further, the ability to measure reporting quality over time by research stakeholders (institutions, publishers, funding bodies) may support targeted research improvement efforts.

Assessing RoB reporting is time-consuming and typically requires two independent reviewers to read through a publication in detail, normally as part of a SR project. To assess RoB across the wider literature, our research group have developed and validated automated regex-based tools to identify RoB measures from the full-text of publications (Bahor et al., 2017). The current sensitivity of the RoB regex tool in identifying random allocation to group, blinded outcome assessment, and the presence of a sample size calculation is above 0.8 (i.e., the tool identifies 80% of publications which report the item). Regexes for blinded outcome assessment and sample size calculations are highly specific (>0.90) meaning that the tools correctly classify over 90% of studies as not reporting the RoB item. The randomisation regex is less specific, with a score between 0.62-0.91. The RoB regex tools face similar limitations to the MIO regex dictionaries in that the regex patterns may match irrelevant words and phrases within the publication. For example, the randomisation regex may match a “randomised controlled trial” listed in the reference section. More recently, we developed two additional regexes to identify where a publication reports approval by an animal welfare committee and a conflict

of interest statement. These have not been extensively validated, so performance remains unclear.

To support automated quality assessment (Chapter 2), I aimed to improve upon and validate regex-based approaches to identify RoB reporting. I tested methods to remove extraneous text (e.g., the introduction and reference sections) from publications and using previously annotated RoB datasets to maximise both sensitivity and specificity.

3.3 Methods

3.3.1 Evaluating tools to remove extraneous text from publications

To improve the specificity of regex approaches for RoB extraction, I evaluated two methods to remove extraneous text from the background and reference sections of publications.

GROBID (Lopez 2008) is a freely available tool for converting PDFs to Extensible Markup Language (XML) format and is available via a Python programming client. XML format is a preferred format for many text-mining and machine learning applications as these files are highly structured and subsections of a document are easily distinguished. Conversion from PDF to XML is unlikely to be perfect in every case, especially if the PDF is in an older format.

I also developed a novel approach to remove extraneous text by modifying how the AutoAnnotation R package cleans text files. The package converts PDF files to plain text before counting the number of regex matches. I wrote R code to capture and remove text from the Background/Introduction section of the text up to the Methods or Results section. To optimise the code for best performance, I added alternative wordings and capitalisations (e.g. “Methods” OR “Methodology” OR “Experimental Procedures” OR “Experimental procedures” OR “Experimental

protocol" OR "Materials and methods") and specified that the headings must be followed by a carriage return and new line (to avoid mistakenly matching with the words in the main body of text). I wrote similar code to remove text below the header for the reference section of a publication. The modified R package with options to remove the background and reference sections is now available on Github (Appendix C3.1).

To determine the usefulness of these approaches, I used RoB annotations from a previous preclinical AD SR (Egan et al., 2016). Working in collaboration with Qianying Wang, PDFs from included publications were converted to XML format using GROBID and the Background and Reference sections of publications were removed. Using the AutoAnnotation package, converted XMLs were converted to plain text before running the validated RoB regex patterns against each of them. For comparison, regexes were also used on an additional set of duplicate PDFs which had not undergone any additional conversion (the standard practice within our group at the time).

3.3.2 Improving RoB regex performance using annotated datasets

In collaboration with Zsanett Bahor, we added functionality to extract the matching strings identified by the AutoAnnotation R package (a link to the latest version of the R package with this functionality is provided Appendix C3.1). This enables us to identify and understand discrepancies between the regex and human decisions. For example, where the regex results suggest that a RoB item is present (where the regex pattern has at least one match within the text), but the human has annotated the item as not reported.

Following this, I utilised a previously annotated dataset from a reporting quality project assessing *in vivo* research studies against the ARRIVE guidelines (Hair et al., 2019). This dataset was particularly useful as reviewers had been asked to note the

sentence from the publication which supported that a RoB item was reported e.g. “animals were randomised to groups”. Using the tool developed to extract matching strings, I was able to see where the regex was picking up strings incorrectly and work towards improving specificity. Where the regex did not detect the RoB measure, I could use the copied sentence from the publication to understand how the human had come to the decision, so I could add in any additional terms of relevance into the RoB regexes. I then compared the performance of the original RoB regexes and the new regexes in the AD SR dataset.

3.3.3 Creating AD-specific MIO regular expression dictionaries

I created regex dictionaries for AD transgenic models, therapeutic interventions, and outcome measures (as described in Chapter 2).

Later, through experience reviewing publications, I came across punctuation and further subtle differences in the way different AD models were described. For feasibility, I focused on optimising a subset of commonly used models of most relevance to my SR projects. I aimed to improve the likelihood that each one would identify most relevant studies using that model i.e., to enhance sensitivity.

To validate MIO extraction, I focussed solely on the extraction of outcome measure and model type. My SR projects (described in Chapters 5 and 6) are focused on commonly used transgenic AD models and specific outcome measures, providing ample data to validate these approaches. In contrast, there were very few publications identified in review with the same intervention, meaning that I could not reliably validate any intervention regex dictionaries.

3.3.4 Evaluating outcome regex matches in abstract vs full-text

Using annotated data from my open field test (OFT) and *in vitro* electrophysiology SRs (Chapter 5/6), I evaluated how often publications reported this outcome

measure in the abstract versus the full-text of included publications. Of note, to obtain relevant datasets of a feasible size and to prioritise publications more likely to contain data from outcomes of relevance, these datasets had been selected based on a criterion of more than three OFT or electrophysiology regex matches within the full-text. All included publications therefore had had at least four matches. I assessed the distribution of matches within included papers to have some indication of whether this criterion was too stringent.

3.3.5 Evaluating model regexes matches in abstract vs full-text

Using annotated modelling data from my open field test SR (see Chapter 6), I tested the sensitivity of selected modelling regexes based on at least one regex match in the full-text or abstract. Assessing specificity is complex due to the nature of many preclinical studies, which may use several different models for different experiments. In some cases, I would extract data using one transgenic model for the review, while other models were mentioned throughout the paper and used to assess other outcomes. To gain some understanding of how specific a regex might be, I evaluated how many other matches there were across the dataset. I also assessed the distribution of regex matches in publications with correctly identified models to determine if there was a "cut off" value indicating that a model is used within the study. I also tested how sensitivity and potential cut-off values change when using the modified AutoAnnotation function to remove extraneous text.

3.4 Results

3.4.1 Removing extraneous full-text

RoB regex performance using the original AutoAnnotation function, with GROBID PDF conversion, and with the modified AutoAnnotation function is shown in Tables

3.1-3.3. Removing irrelevant text from publications using GROBID was successful in improving the specificity of RoB regex approaches. Randomisation regex specificity in this dataset improved from 0.579 to 0.656, while other regexes displayed more modest improvements. However, this was accompanied by a drop in sensitivity. Randomisation regex sensitivity dropped from 0.963 to 0.927, blinded outcome assessment dropped from 0.862 to 0.850, and approval by an animal welfare committee dropped from 0.675 to 0.621. Conflict of interest regex detection was most severely impaired by GROBID conversion (0.078, 4/55 publications); possibly indicating that information at the end of the publication might, in addition to the reference section, have also been removed, likely due to incorrect XML conversion. Further, when looking at the text files resulting from the GROBID conversion (XML -> PDF -> text), it was clear that chunks of text had been removed unintentionally from various parts of the publications. Using the modified AutoAnnotation R function led to similar improvements in specificity (Randomisation rising from 0.569 to 0.625; Blinding increasing from 0.914 to 0.923), but without substantial drops in sensitivity. Only the regex for conflict of interest statement showed a reduction from 0.784 sensitivity to 0.765 (the statement was no longer identified in 1 publication versus original regex function). It was not possible to assess the performance of the same size calculation regex, as there were no publications which reported the use of a sample size calculation in this dataset.

	Sensitivity	Specificity	True +	True -	False +	False -
Blinding	0.86	0.91	75	247	23	12
Approval by animal welfare committee	0.67	0.85	139	129	22	67
Conflict of interest	0.78	0.99	40	302	4	11
Randomisation	0.96	0.58	53	175	127	2
Sample size calculation	?	0.96	0	344	13	0

Table 3.1: RoB regex performance using original AutoAnnotation function

	Sensitivity	Specificity	True +	True -	False +	False -
Blinding	0.86	0.93	74	251	19	13
Approval by animal welfare committee	0.62	0.86	128	130	21	78
Conflict of interest	0.08	1	4	306	0	47
Randomisation	0.93	0.66	51	198	104	4
Sample size calculation	?	0.96	0	344	13	0

Table 3.2: RoB regex performance using original AutoAnnotation function with GROBID PDF to XML conversion

	Sensitivity	Specificity	True +	True -	False +	False -
Blinding	0.86	0.93	75	250	20	12
Approval by animal welfare committee	0.67	0.85	139	129	22	67
Conflict of interest	0.76	0.99	39	302	4	12
Randomisation	0.96	0.63	53	189	113	2
Sample size calculation	?	0.96	0	344	13	0

Table 3.3: RoB regex performance using modified AutoAnnotation function

3.4.2 Improving the RoB regexes

Using annotated data from a previous publication (Hair et al., 2019), I modified the current regexes to improve their accuracy (see Appendix C3.1 for latest RoB regexes). Performance using the original function and modified Auto Annotation function is shown in Tables 3.4 and 3.5. By adding additional terms from previously annotated data, I was able to improve the sensitivity of most RoB regexes. Further, by examining the regexes which matched incorrectly, I was able to notice small errors where the regex matched irrelevant text, impacting the specificity. Sensitivity dropped slightly for compliance with animal welfare statements and conflict of interest statements using the modified AutoAnnotation function.

	Sensitivity	Specificity	True +	True -	False +	False -
Blinding	0.940	0.908	78	188	19	5
Approval by animal welfare committee	0.692	0.886	126	93	12	56
Conflict of interest	0.864	0.996	38	245	1	6
Randomisation	0.898	0.797	44	192	49	5
Sample size calculation	?	0.955	0	277	13	0

Table 3.4: New RoB regex performance

	Sensitivity	Specificity	True +	True -	False +	False -
Blinding	0.940	0.918	78	190	17	5
Approval by animal welfare committee	0.687	0.886	125	93	12	57
Conflict of interest	0.841	0.996	37	245	1	7
Randomisation	0.898	0.813	44	196	45	5
Sample size calculation	?	0.955	0	277	13	0

Table 3.5: New RoB regex performance using modified AutoAnnotation function

3.4.3 Regex approaches to identify outcome

Across the 287 publications included in the OFT dataset, only 52/287 (18.4%) had a regex match for OFT in the title or abstract. Within the full-text, there were between 4 and 40 regex matches, with a median of 8 matches. The distribution of matches is shown in Figure 3.2.

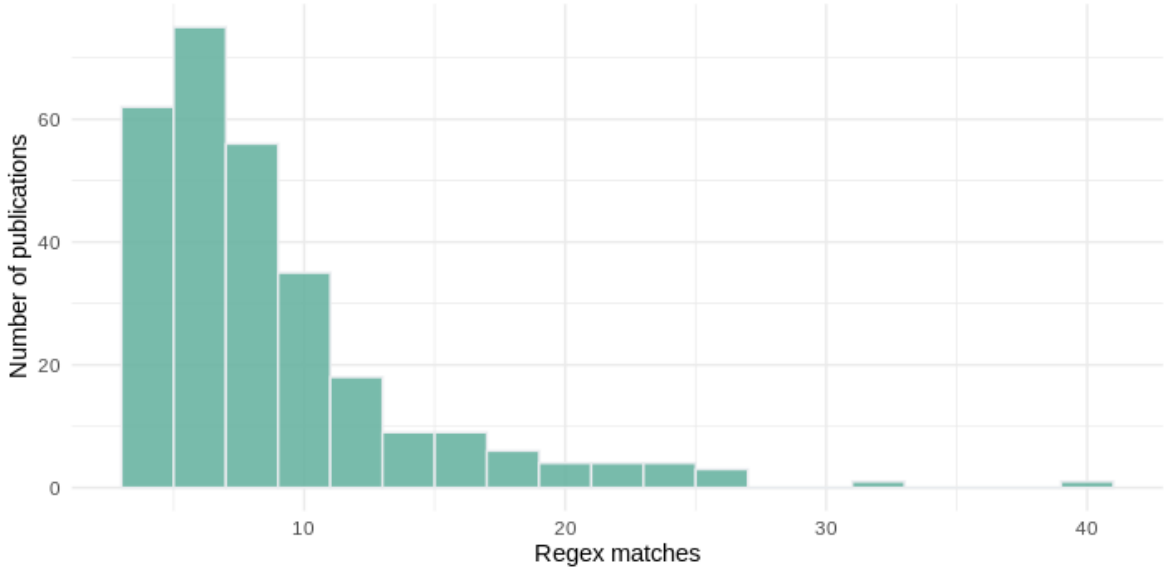


Figure 3.2: Outcome regex match frequency in OFT dataset

For the electrophysiology dataset, 51/166 (30.7%) of publications had a regex match in the title or abstract. Within the full-text, there were between 4 and 142 regex matches identified in included publications, with a median of 18 matches. The distribution of electrophysiology regex matches is shown in Figure 3.3.

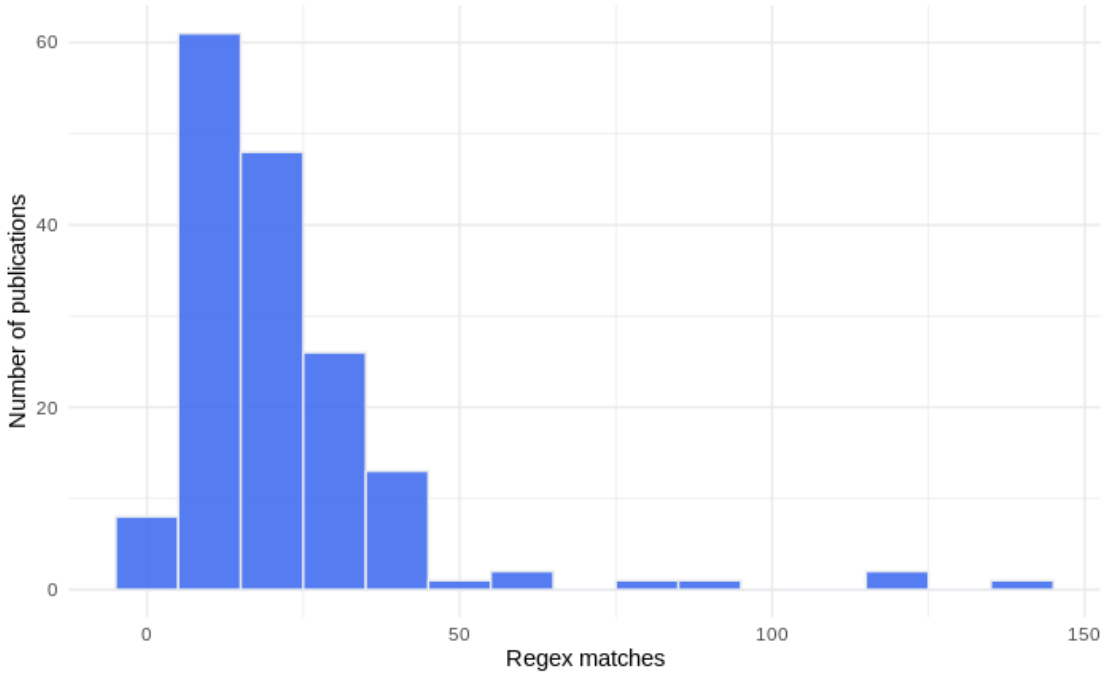


Figure 3.3: Outcome regex match frequency in electrophysiology dataset

3.4.4 Regex approaches to identify animal model

The sensitivity for each model regex using either the original AutoAnnotation function, modified function, or using just the title/abstract of the publication is shown in Table 3.6. There were fewer regex matches in the title/abstract for publications in most models. In the APPSwe/PSEN1de9 model, only 40% of publications had a APPSwe/PSEN1de9 regex match in the title or abstract versus 90.7% of publications with regex matches in the full-text using the original AutoAnnotation function. Most model regexes had high sensitivity i.e., publications in that model had at least one regex match. The APPPS1 and APP23 regexes may

require additional work to identify why sensitivity is not optimal (25.9% and 60% respectively using the original AutoAnnotation function). In most cases, the modified AutoAnnotation function, which selectively removes reference and background sections of the full-text, had similar sensitivity to the original function, but there were some instances where it reduced slightly (3xTg and J20, and APPSwe/PS1de9 models). This may be due to a description of the transgenic model in the background/introduction sections being different to the description in the methods/results sections, or because the code has incorrectly removed areas of useful text.

Although specificity cannot be accurately measured here, I have also presented the number of matches to publications using other models. Using the modified AutoAnnotation function produced less regex matches in other publications for most models. This may indicate some improvements in sensitivity by removing extraneous text. The use of modelling regexes using either the modified or original AutoAnnotation function has the potential to reduce workload, as indicated by the number of non-matching publications. If investigators performing a SR wanted to assess publications in a specific transgenic model, most model regexes only match to a fraction of the remaining dataset and would therefore exclude a large proportion of publications from consideration. Regex matches in the title / abstract were highly specific, and few matches were identified in publication which did not use the specified model.

The distribution of the frequency of regex matches within publications using each model (the original AutoAnnotation function and the modified function) is shown in Figures 3.4 and 3.5. I have omitted the APPS1 and APP23 regexes from this analysis due to poor regex sensitivity and several publications with 0 regex matches. Median frequencies across models ranged between 6-76 using the original AutoAnnotation function and 4-75.5 using the modified AutoAnnotation function. There was substantial variation across model regexes, with the APPSwe/PSEN1de9 model

matching fewer times within included publications. From my experience conducting SRs of the literature using transgenic AD models, I have observed that often this model is abbreviated as APP/PS1 and the full description may only be mentioned once or twice. Looking at publications across all models, the minimum match frequency was often only one or two. Therefore, for this model approach to be as sensitive as possible, any number of matches may indicate that that animal model is used within the paper, and it is therefore difficult to ascertain what a good cut off value would be to ensure maximal sensitivity and specificity.

Regex	Model	True +	False -	Other matches	Non-matches	Sensitivity
Original	3xTg	47	0	78	162	100.0%
Modified		46	1	76	164	97.8%
TIAB		46	1	10	230	97.8%
Original	APP23	6	4	63	214	60.0%
Modified		6	4	58	219	60.0%
TIAB		6	4	0	277	60.0%
Original	Tg2576	32	1	107	147	97.0%
Modified		32	1	103	151	97.0%
TIAB		27	6	8	246	81.8%
Original	TgCRND8	14	0	33	240	100.0%
Modified		14	0	32	241	100.0%
TIAB		11	3	0	273	71.4%
Original	J20	27	0	16	244	100.0%
Modified		26	4	16	241	96.3%
TIAB		10	17	0	260	37.0%
Original	5XFAD	29	1	15	242	96.7%
Modified		29	1	15	242	96.7%
TIAB		26	4	0	257	86.7%
Original	APPSwe/PS1de9	59	6	59	163	90.1%
Modified		57	8	58	164	87.7%
TIAB		26	39	5	217	40.0%
Original	APPPS1	7	20	8	252	25.9%
Modified		7	20	8	252	25.9%
TIAB		3	24	0	260	2.5%

Table 3.6: Performance of modelling regexes in OFT SR dataset.

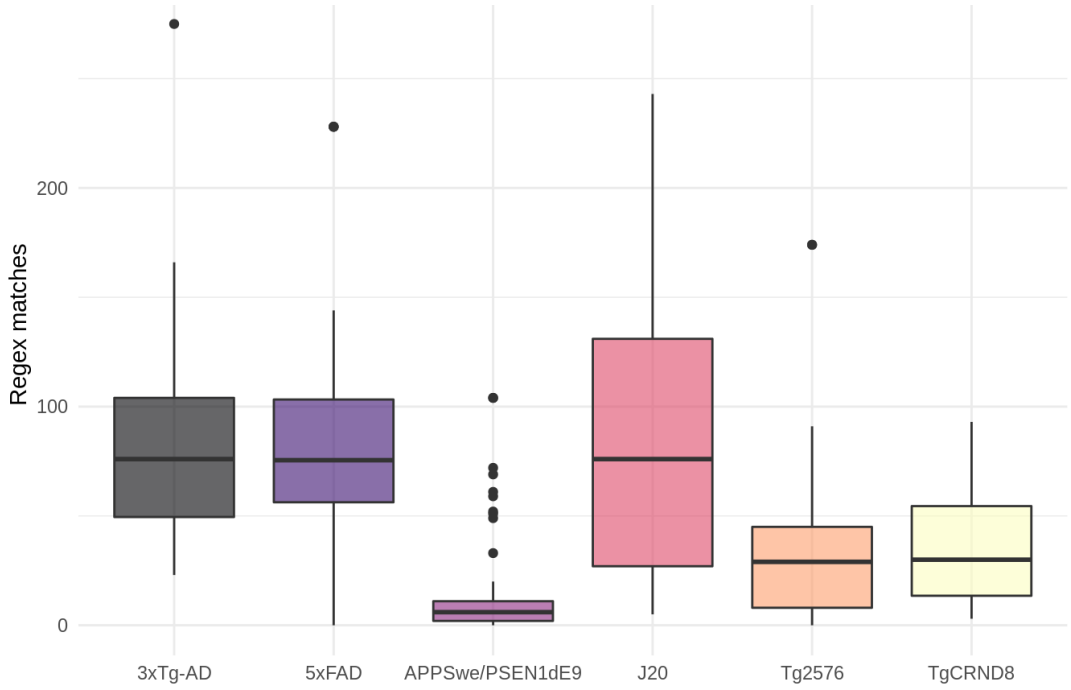


Figure 3.4: Distribution of regex matches across model regexes

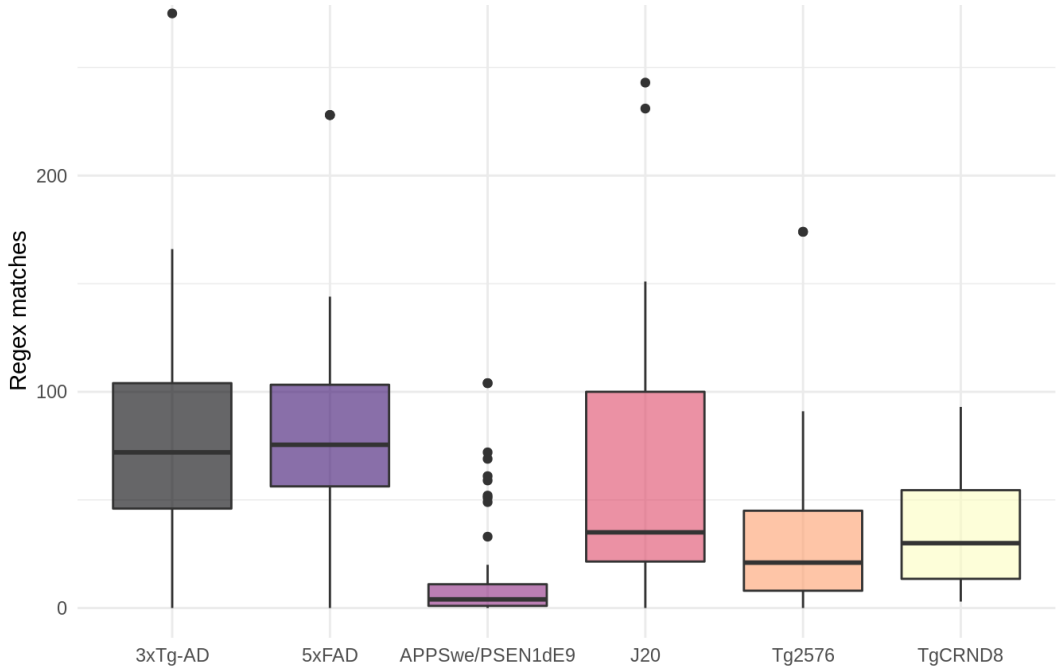


Figure 3.5: Distribution of regex matches across model regexes (modified AutoAnnotation)

3.5 Discussion

In this chapter, I have developed and evaluated regex-based approaches for automated PICO extraction and improved upon current methods for RoB assessment.

3.5.1 Improving RoB regexes

Both methodologies (GROBID conversion and the modified AutoAnnotation tool) were successful in improving the specificity of RoB regexes. However, the GROBID approach was error-prone and, in some cases, substantially reduced sensitivity. Using the modified AutoAnnotation function resulted in modest improvements in specificity, without affecting sensitivity.

Modifying the function using annotated RoB data from a previous project led to marked improvements in the performance of RoB regexes. The modified AutoAnnotation function led to further improvements in specificity, however, there was a small drop in sensitivity. By using more annotated datasets, ideally with sentences containing the reporting RoB items extracted from the full-text, the RoB regex tools can likely be improved further. Overall, these findings suggest that this may be the preferred approach over either GROBID conversion or the modified AutoAnnotation function, which can each lead to a reduction in regex sensitivity. The latest version of the RoB regex tools are available as part of an R package (Appendix C3.2) developed in collaboration with Jing Liao and Zsanett Bahor.

3.5.2 Reporting of outcome measures in the title or abstract

For MIO extraction, I focussed solely on the extraction of outcome measure and model type. My SR projects (described later in this thesis) are focused on a list of commonly used transgenic AD models and specific outcome measures, providing ample data to validate these approaches. In contrast, there were very few

publications identified in review with the same intervention, meaning that I could not reliably validate any intervention regex dictionaries. The outcome measure regexes (OFT, *in vitro* electrophysiology) often did not match anything in the title or abstract of included publications. This was more apparent for the OFT (18.4% had a match) versus electrophysiology studies (31.1% had a match). This has important ramifications for preclinical SRs focussed on specific outcome measures. I used a regex-based approach to identify outcomes for my SR projects, but searched databases only on the outcome measure, and it is likely that I would have missed a substantial proportion of the literature.

3.5.3 Outcome regex match frequency

Included publications for each SR were selected based on over 3 outcome regex matches. For both outcomes, the median number of matches was higher, indicating that most publications did mention the outcome measures more than a few times. However, in future work, I intend to examine publications which matched each outcome regex between 1-3 times by extracting the matching text from within each publication.

3.5.4 Reporting of animal model in the title or abstract

Model regexes have reduced sensitivity in detecting the model in the title/abstract of publications versus the full-text. However, when a model was specified in the title/abstract, it was highly likely that the publication did use that model in experiments. Using model regexes in the full-text was highly sensitive for most models. Although specificity was likely not optimal, as many matched to several other publications, this approach could at least reduce the time required to go through and read additional publications. Therefore, it could still be a useful approach to flag studies that mention specific models.

3.5.5 Model regex match frequency

I was not able to establish a useful cut-off value which indicated that a model was definitely used in experiments, as the distribution often ranged from just 1 or 2 regex matches to hundreds of regex matches. Instead, the modelling regexes are perhaps most useful as a signposting tool to indicate that a publication may use a given model.

3.5.6 Limitations

One of my aims for this chapter was to improve PICO element extraction from publications. However, I focussed only on population (transgenic model) and outcome reporting. I did not have enough publications focused on specific interventions in either of my reviews to assess the validity of any intervention regex. Further, to my knowledge, there have not yet been any attempts to extract comparator information from preclinical studies. The utility of this may be limited, as there would be no way of knowing which comparator matched to which outcome measure. However, extracting which papers reported using wild-type littermates versus other types of controls may be of interest.

The AD-SOLES pilot web application (Chapter 2) allows users to select publication subsets by filtering for specific models, outcomes, and treatments. However, as each publication often contains several experiments, it is unclear which treatment was tested in which transgenic model and which outcome measure was assessed. Future work may evaluate how close each regex match appears in a publication, with the hypothesis that the therapeutic intervention may be mentioned within the text beside the behavioural test used to measure its effect.

The strength of automated tools will always be limited by how clearly the details are presented in the paper. Throughout my SR projects, I noted that the model was often unclear. Often, I had to check specific mutations that were mentioned and

decide which model the publication was most likely to describe. Also at times, the original publication describing the development of the model might be cited, assisting this decision. However, until models are described in sufficient detail and in a consistent manner throughout the literature, it will be challenging to improve the accuracy of automated tools.

3.5.7 Future perspectives

As many publishers and databases now have the option to download full-texts in XML format, future work could explore the feasibility of using regex tools in specified sections (methods/results) of these files compared to PDF to text conversion approaches.

Importantly, the closed infrastructure which surrounds research publications prevents us from deriving their maximum potential. As set out in the FAIR principles for data management and stewardship, “research objects should be findable, accessible, interoperable, and reusable” (Wilkinson et al., 2016). A lack of sufficient meta-data to classify a study accurately (e.g., unstructured abstracts which do not mention key information about the study), inaccessibility (e.g., no open-access PDF available), and differences in the way vital study information is described (e.g., variation in model and drug nomenclature) complicates automated evidence synthesis approaches. The Resource Identification Initiative (RRID) promotes the use of unique identifiers to enable researchers to cite the exact methods used in their experiments (Bandrowski & Martone, 2016). For example, each transgenic model should (and may already) have a unique RRID number. If widely adopted by the AD research community, this would transform our ability to detect models, treatments, and outcomes across publications.

3.6 Conclusion

In this chapter, I describe my work to improve current regex-based approaches to extract PICO elements from publications and to automatically assess RoB. Findings from this work will influence future iterations of the AD-SOLES project, to signpost users to publications that are likely to contain experiments in a given model or which measure effect using a specific outcome measure. Although this automated classification will not be able to determine whether a model or outcome is used within a publication, this approach could reduce time spent looking through irrelevant publications. Using human annotated data from a previous project, I improved the performance of regex tools to automate RoB assessment.

CHAPTER 4: DEVELOPING AN AUTOMATED TOOL TO REMOVE DUPLICATE CITATIONS FROM LARGE-SCALE SYSTEMATIC SEARCH DATASETS

4.1 Chapter Introduction

In this chapter, I discuss current approaches to deduplicate SRs and describe how I developed and validated a new, automated tool I developed specifically for the deduplication of systematic searches for preclinical SRs.

4.2 Background

4.2.1 What are duplicate publications?

Researchers performing a SR typically search across multiple biomedical databases to collect as many relevant citations as possible (Paul, Michael, & Daniel, 2015). This process can introduce a substantial number of duplicate citations (handQi et al., 2013). For example, overlap between EMBASE and PubMed is estimated to be as much as 79% (Royle & Milne, 2003). To complicate matters further, although publication of the same article in more than one journal is widely considered to be unethical (at least under most circumstances), I have identified many examples of this while conducting SRs. In fact, six different patterns of duplicate publication have been identified (von Elm, Poglia, Walder, & Tramèr, 2004) ranging from a direct "copy" of an article to so-called "salami" publications which slice up the data from one dataset into many resulting publications in an inappropriate manner (Abraham, 2000). Such practices threaten scientific integrity and inflate redundancy in the literature (Huston & Moher, 1996). Different approaches may be required to tackle the distinct forms of "duplicate publication".

Effective duplicate removal is an essential, if underappreciated, part of the data collection process of SRs (Qi, Bai, Yang, & Ren, 2013). If duplicate citations are not removed effectively, reviewers can waste time screening the same citations for

inclusion, and run the risk of accidentally including the same paper more than once in their meta-analyses, leading to inaccurate conclusions (Tramer, Reynolds, Moore, & McQuay, 1997). False positives (incorrect removal of citations which are not duplicates) can be just as problematic (Jiang, Yu et al., 2014; Kwon, Lemieux, McTavish, & Wathen, 2015) and reduce the accuracy and reproducibility of SRs.

Here I consider the challenge of bibliographic duplicate detection – where the same publication in the same journal is retrieved from several biomedical databases. Current and common approaches to deduplication for SRs are summarised in Table 4.1.

Tool	Description	Resource required	Accessibility	Performance
Endnote (Hupe, 2019)	Reference manager	**Medium** (requires some manual effort to improve sensitivity)	**Medium** (requires a paid subscription)	**Low – Medium** (user-dependent) (Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T., 2016; Kwon et al., 2015; Rathbone, Carter, Hoffmann, & Glasziou, 2015)
SRA-DM (Rathbone et al., 2015)	Web/desktop application	Low	High	High
Revtools (Westgate, 2019)	R package	**Medium** (users often need to set parameters for deduplication within the function)	**Medium** (some R knowledge required)	Unknown
Metta (Smalheiser et al., 2014)	Cross database search engine	Low	**Medium** (not openly accessible)	**High** (Jiang, Y. et al., 2014)
Zotero (Mueen	Reference manager	**Medium** (manual	High	Unknown

Ahmed & Al Dhubaib, 2011)		merging required)		
Mendeley (Zaugg, West, Tateishi, & Randall, 2011)	Reference manager	**Medium** (manual merging required)	High	**High** (Kwon et al., 2015)
Hand-searching	Manual	High	Low	**High** (Qi et al., 2013)

Table 4.1: Deduplication tools and approaches for SRs

SR findings often inform clinical practice. In recent years, largely in response to discrepancies between findings in laboratory research and clinical trial results, researchers have begun to apply SR methodologies to summarise preclinical evidence from animal and cell models of disease (de Vries et al., 2014; Sena et al., 2014). When considering different tools to identify duplicates from preclinical systematic search datasets, we must consider what type of citation data the tool was designed to deduplicate. For example, the databases supported by Metta are highly specific to clinical research and do not support search engines routinely used for preclinical reviews such as Web of Science. Furthermore, the type and extent of duplicate publications may differ in the preclinical literature – an author may publish a higher number of similar papers in a short space of time, or there may be less bibliometric information available for studies published in lesser known (and less frequently indexed) journals. Our group frequently retrieves tens of thousands of potentially relevant citations for a preclinical SR. Tools should therefore be evaluated on comparatively large datasets to determine the magnitude of gains and losses on that scale (e.g. how many duplicate citations a tool is likely to remove correctly). Previous evaluations of duplicate removal tools have used relatively small (<5,000 citations) systematic search datasets primarily representing clinical research citations (handQi et al., 2013; Kwon et al., 2015; Rathbone et al., 2015).

4.2.2 Current methods of duplicate removal

Researchers often use citation managers to remove duplicates, as these are easy to use and straightforward to integrate into typical systematic search methods. Among them, Endnote is one of the most established (Lorenzetti & Ghali, 2013). Endnote's "find duplicates" feature automatically detects citations matching on Author, Year and Title by default. Users can also adjust the match criteria within Endnote's settings (i.e. match on Title and Journal) to identify additional duplicate records. The requirement for a 100% match to identify duplicates, however, results in many records being missed. Small differences in the way the Titles, Authors, and Journals are represented are extremely common. Deduplication might be simplified through the use of unique identifiers for journal articles such as PubMed IDs (PMIDs) or digital object identifiers (DOIs). However, Endnote does not provide an option within their deduplication settings to match citations based on DOIs, PMIDs, Accession Numbers, or URLs. Matching is further complicated by indexing differences in the formatting of page ranges, with some biomedical databases adopting a longer form (1234-1235) and some a shorter form (1234-5); although an import filter has been developed to address this issue in Endnote (Bramer, Wichor M. et al., 2016).

Endnote's auto-deduplication feature is an attractive option due to its simplicity, yet there is a wealth of evidence to suggest it is an imperfect solution, as it fails to identify more duplicates (higher number of false negatives) and removes more citations incorrectly (higher number of false positives) than other citation managers (Kwon et al., 2015). Moreover, our prior experience of using Endnote is that many duplicates remain in large datasets even after extensive deduplication using a combination of automated and user-configured methods.

Many citation managers, including Endnote, are proprietary software which restricts their accessibility, prevents interoperability, and limits transparency about

how their underlying duplicate detection process works. Increasingly, freely available open-source citation managers such as Zotero and Mendeley have gained popularity. Both have integrated deduplication tools which match citations automatically, then require users to manually select citations to merge within each matching group.

Several other tools for duplicate removal have emerged in recent years, either as stand-alone tools or as part of alternative workflows (which may bypass the need for traditional citation managers). The "Systematic Review Assistant" (SRA) is a suite of free, open-source systematic review tools developed by researchers at Bond University. Their "deduplication module" (SRA-DM) has a user-friendly interface in which users can upload a search file in various formats and perform automated duplicate removal in a few clicks. SRA-DM has been shown to identify substantially more duplicates than Endnote (Jiang, Y. et al., 2014; Rathbone et al., 2015). Another option is the metasearch engine Metta, which automatically removes duplicate citations appearing across 5 medical databases including PubMed, EMBASE, CINAHL, PsycINFO and Cochrane Central Register. De-duplication is also possible using Revtools, an R package. Of course, manual deduplication is strongly advised to complement these automated approaches (handQi et al., 2013), but this is time consuming and can lead to errors (Kwon et al., 2015).

4.2.3 Deduplication tools to support "Living" or automated reviews

Increasingly, meta-researchers are aspiring to provide automated or "living" systematic reviews (Elliott et al., 2014), producing real-time summaries of a domain including the most recent research findings. To enable such summaries, we need automation tools at each stage that are reliable and require minimal manual intervention. Where review teams are large, as is the case in crowdsourced reviews, the risk of duplicate studies being retained is likely higher. Sensitivity of a deduplication tool (ability to detect duplicates) is therefore of paramount

importance, since several reviewers could extract information from a given paper, unaware that others were also doing so. Furthermore, if machine learning approaches are used to select included studies, duplicate publications present in the training data may reduce the performance of classifiers.

Deduplication tools should be interoperable and easily integrated into automated workflows. Tools with a programmatic component are likely superior in this respect because once they have been configured, they may be implemented in a data pipeline without manual intervention. Depending on the project goals, it may be useful to have some control over the tool's duplicate removal logic. For instance, if two records are identified to be duplicates of each other, which record should be retained? It may be useful to configure the tool to retain the existing version of a citation when a new, matching citation is identified in an updated search, so that existing annotations and data extractions can be retained. This approach could also be used in more conventional systematic review updates, often occurring after many years (Bashir et al., 2018) and often involving significant overlap between systematic search dates to prevent missing relevant studies. Alternatively, researchers may wish to preferentially retain the newer citation, which may be more complete and may contain more accurate meta-data.

I developed the Automated Systematic Search Deduplicator (ASySD) to identify and remove bibliographic duplicates from preclinical systematic review searches. The tool allows users to label which reviews should be preferentially maintained (e.g. older citations), and can be accessed either through a web application or integrated programmatically with automated workflows via an R package. I critically evaluated ASySD in comparison with two user-friendly, low effort automated tools - Endnote's automated duplicate removal and Bond University's SRA-DM.

4.3 Methods

Prior to performance evaluation I registered a protocol describing our methods on the Open Science Framework (see Appendix C4.1).

4.3.1 Definition of "Duplicate citations"

I define bibliographic duplicates as the presence of two or more citations representing the same publication within an aggregated systematic review search result, even where those citations differ subtly in recorded details such as author(s), title, journal pagination, issue number or volume. If the same study is published in two separate journals, I did not consider this a duplicate citation for these purposes. Similarly, sets of conference abstracts, preprints and journal articles which describe the same research are not classed as duplicate citations.

4.3.2 Tool development and functionality

I developed ASySD in the R programming language. To improve the chance of detecting duplicate citations, data undergoes several cleaning and formatting steps. This includes renaming missing or anonymous Authors as "Unknown", harmonising differences in DOI format, removing punctuation, and making all citation information upper case.

Using the RecordLinkage R package (Borg, 2010), I applied blocking criteria (fields which must be a 100% match) to identify possible duplicate pairs. These criteria were largely based on guidance to systematically identify all possible duplicates using Endnote's manual 100% match filters (Bramer, Wichor M. et al., 2016). Blocking criteria (see Table 4.2) were applied in four separate rounds because of the extensive memory requirements needed to perform these operations on large datasets in R. However, matches identified within any of the rounds were considered a possible duplicate pair.

Order	Blocking criteria (100% match on specified fields)
Round 1	(**Title** AND **Pages**) OR (Title AND Author) OR (Title AND Abstract) OR DOI
Round 2	(Author AND Year AND Pages) OR (Journal AND Volume AND Pages) OR (ISBN AND Volume AND Pages) (Title AND ISBN)
Round 3	(**Year** AND **Pages** AND **Volume**) OR (**Year** AND **Issue** AND **Volume**) OR (Year AND Pages AND Issue)
Round 4	(**Author** AND **Year**) OR (**Title** AND **Year**) OR (**Title** AND **Volume**) OR (**Title** AND **Journal**)

Table 4.2: Blocking criteria to identify potential pairs of matching publications

Most pairs identified with blocking criteria are not true duplicates, and further comparisons are needed to ascertain duplicate status. To compare the overall similarity of a matching pair, I also calculate string comparisons across all relevant fields (Title, Year, Journal, ISBN, Abstract, DOI, Issue, Pages, and Volume) using the RecordLinkage package. Using a heuristic approach, I developed and applied additional match filters based on string comparison match strength (a numerical value between 0 and 1) to optimise performance and prevent the deletion of citations which were not duplicates. During development, I used three existing CAMARADES systematic review search results with labelled duplicates (Neuropathic Pain (Currie et al., 2018), Antioxidants (McCann, 2018), and Epilepsy (Simonato et al., 2017) to iteratively validate and adjust the match filters to improve the performance of the tool.

Once ASySD has identified all matching citations, one citation is removed from each pair. Firstly, citations which do not contain abstracts are preferentially removed. Where a newer version of a citation exists (e.g. e-publication date versus

publication date), I will preferentially retain the most up-to-date version. If neither of these rules apply (e.g. both citations contain abstract text, and have the same year of publication), then the second listed citation in each pair is removed. Where there are more than two duplicates, the code logic ensures that only one is kept from within each duplicate set. There is an option for users to set a preference for citations to be retained in the dataset using a "Label" field. If specified, duplicates are ordered so that these citations are always the first citation in each pair and therefore retained.

Citation pairs which fall short of the additional match filters but still have high string comparison scores are retained for manual deduplication – where users can manually review these matches and select which (if any) citation of the two they would like to remove from the search.

The underlying code for ASySD is open-source and available on Github, where it is also available to download as an R package (Appendix C4.2) To ensure accessibility, I have also created a user-friendly web application built using R Shiny (Appendix C4.3) Users can upload a file with search returns (e.g. Endnote .xml, .csv, or .txt file), click a button to run the deduplication procedure, complete any additional manual deduplication within the application (if required), and download the results as a .csv file or a tab delimited .txt file (formatted for importing into Endnote). For transparency, there is the option to download a file with all potentially matching pairs side-by-side (from initial blocking criteria) and to download all matching pairs after the additional filters were applied. The code underlying the Shiny web application is also available on Github (Appendix C4.4).

4.3.3 Gold-standard systematic search datasets

I assessed the performance of automated deduplication tools on five test datasets of varying sizes from systematic review searches (Table 4.3). For each dataset,

duplicate citations had been removed in Endnote using a combination of automated deduplication functions, changing field parameters to identify all citations which match on certain field e.g. “Title”, and manual checking. Citations which had been removed by the human reviewer were reinstated and labelled as duplicates. I obtained three systematic search datasets from external sources, described below. I also used two datasets curated as part of ongoing in-house projects, a systematic review of systematic reviews of animal models of human disease (SRSR), and a systematic review of animal models of depression. Importantly, none of these datasets had been used in the development of the tool. To assess the time taken to perform “gold-standard” deduplication, I measured the time taken to deduplicate the SRSR dataset. To identify duplicates, I imported the systematic search into Endnote and followed recommended guidance (Bramer, Wichor M. et al., 2016) to systematically identify all duplicate citations in the dataset using a range of different matching field parameters e.g. matching on “Author” and “Year”.

Dataset description	**Databases searched**	**Citations obtained**	**Duplicates removed**	**Citations remaining**
Diabetes dataset*: Antidiabetics in animal models of atherosclerosis (SYRCLE, Radboud University) (Wever, Ranis, Hooijmans, & Riksen, 2018)*	Pubmed, EMBASE	1,845	896	949
Neuroimaging dataset: *Epigenetic neuroimaging (MRC Centre for Reproductive Health, University of Edinburgh) (Wheater et al., 2020) (Preclinical (in vivo) and clinical data included in review)*	SCOPUS, EMBASE, Medline, Web of Science,	3,438	1,280	2,158
Cardiac dataset: *Efficacy of cardiac ischemic preconditioning in animal models (SYRCLE, Radboud University) (Wever et al., 2015)*	Pubmed, EMBASE	8,948	3,153	5,795

***Depression dataset**: Preclinical animal models of Depression (CAMARADES, University of Edinburgh) (Bannach-Brown, Liao, Wegener, & Macleod, 2016)*	PubMed, EMBASE, Web of Science,	79,880	9,418	70,462
***Systematic review of systematic reviews (SRSR) dataset**: Systematic review of preclinical systematic reviews dataset (CAMARADES, University of Edinburgh) (Hair & McCann, 2020)*	PubMed, EMBASE, Web of Science,	53,001	16778	36223

Table 4.3: Gold standard systematic search datasets

4.3.4 Methods for performance evaluation in testing datasets

To obtain the most up-to-date citation information and ensure all systematic searches for validation have a similar depth of information, I used the "find reference updates" feature in Endnote X9 to retrieve additional information (e.g. DOIs, page numbers, issue numbers, journal volumes).

I compared the performance of the ASySD tool (automated, with no manual input, deduplication mode only), Endnote X9 automatic deduplication, and SRA-DM (Rathbone et al., 2015) on the five gold-standard search datasets. To assess auto-deduplication performance using Endnote X9, I auto-deduplicated citations based on "author", "year" and "title" matching criteria and using the "ignore spacing and punctuation" feature. In SRA-DM, I uploaded XML files of our datasets to the offline version of the tool (as the server has limited capacity for high volume datasets) and chose the automated deduplication option to remove all suspected duplicates. In the ASySD tool, I uploaded citations as an XML file to the web application and ran automated deduplication. Because of memory limitations on the shinyapps.io server, for search results containing over 50,000 citations, I ran the R Shiny application locally in R.

To preferentially retain records which had been labelled as duplicates by the human reviewer (so that I would know that these had been identified as duplicates), I used the "labelled duplicates" feature of ASySD to preferentially remove citations which the human had also removed. Importantly, this process does not affect the accuracy of the tool – only the choice of which citation from each pair is removed. This made the deduplication process of ASySD as similar as possible to that of the human reviewer, to make it simpler to assess the performance of the tool.

Once duplicates were removed using each of the other tools, a "Duplicate ID" was generated for matching sets of duplicates identified by ASySD. This was possible because ASySD allows users to download the Record IDs of matching citation pairs. There should, therefore, for each Duplicate ID be one single citation labelled as "KEEP" and the remainder (one or more duplicate citations) labelled as "REMOVE". I carried out extensive manual checking in MS Excel to interrogate duplicate citations identified by some approaches but missed by others, to ensure that they were indeed duplicates. I manually searched to identify additional studies and corrected the Duplicate ID as appropriate. All data (including the original de-duplicated search datasets, results from each deduplication tool, final manually checked datasets with duplicate IDs, and the R code used to assess performance) is available on the Open Science Framework (link in Appendix C4.5). Once each search file had been corrected, I analysed this final dataset in R to calculate performance.

I reported the performance of each tool by calculating:

- Number of true positives (citations which are duplicates which are correctly removed from the dataset);
- Number of false positives (citations which are not duplicates which are wrongly removed from the dataset);
- Number of true negatives (citations which are not duplicates which correctly remain in the dataset);

- Number of false negatives (citations which are duplicates which remain in the dataset but which should have been removed).
- Precision $= \frac{true\ positive}{true\ positive + false\ positive}$
- Sensitivity (Recall) $= \frac{true\ positive}{true\ positive + false\ negative}$
- F1 score $= 2 \cdot \frac{precision \cdot recall}{precision + recall}$

I also recorded any duplicates found by any of these approaches which had not been identified by humans in our "gold standard" datasets. I also recorded the time taken by each tool to deduplicate each dataset.

4.4 Results

4.4.1 True duplicates identified by any method

Across all datasets, additional duplicates were identified by automated tools which had been missed by the human reviewer(s). Furthermore, a small number of citations had been removed incorrectly by the human reviewer(s). I carefully considered all discrepancies between human reviewers and the automated tools to derive a new "gold standard" annotation against which to compare all approaches.

4.4.2 Diabetes dataset

The Diabetes dataset (N=1,845) had 1,261 duplicate citations (68.3% of total; Table 4.4), of which 896 had been identified by human reviewers in the course of the systematic review, and a further 368 identified by at least one of the automated approaches and later confirmed by human scrutiny. While the sensitivity of the human approach was low, the specificity was high; only three citations were removed which were not duplicates (Table 4.5). Endnote, the SRA-DM, and ASySD

were highly sensitive (sensitivity = 0.966, 0.910, and 0.998 respectively), but SRA-DM had a higher rate of false positives (n=70 citations incorrectly removed). The ASySD tool outperformed all other automated methods in terms of sensitivity (0.998), specificity (1.0), precision (1.0), and F1 score (0.999). Each automated deduplication method took less than 5 minutes to identify and remove duplicates in the diabetes dataset.

Deduplication method	**Duplicate citations removed**	**Citations remaining**
TRUE duplicates (all methods + hand searching)	1261	584
Human	896	949
Endnote (automatic)	1218	627
SRA-DM	1217	628
ASySD	1259	586

Table 4.4: Duplicate citations identified in the diabetes dataset by each method

	TP	**TN**	**FN**	**FP**	**Sens**	**Spec**	**Prec**	**F1**	**Time**
Human	893	581	368	3	0.708	0.995	0.997	0.828	?
Endnote	1,218	584	43	0	0.966	1.0	1.0	0.983	<5 m
SRA-DM	1,147	514	114	70	0.910	0.880	0.942	0.926	<5 m
ASySD	1,259	584	2	0	0.998	1.0	1.0	0.999	<5 m

Table 4.5: Performance of each deduplication tool in the diabetes dataset
TP= true positive, TN = true negative, FN = false negative, FP=false positive, Sens = Sensitivity, Spec = Specificity, Prec = Precision, m = minutes

4.4.3 Neuroimaging Dataset

The Neuroimaging dataset (N = 3,434) had 1293 duplicate citations (37.2% of total; Table 4.6). In this dataset, the human reviewer was highly sensitive and identified the vast majority of duplicate citations (sensitivity = 0.985; Table 4.7). However, a few citations had been removed in error (n=6), and a small number of duplicate citations were missed (n=19). Automated deduplication by Endnote and the SRA-DM was lacking in sensitivity and each missed hundreds of duplicates (n=310 and 243 respectively). The SRA-DM incorrectly removed a substantial number of citations (n=42). The false positives rate of the ASySD (n=4) and Endnote (n=3) were

comparable to human performance. Overall, the ASySD tool outperformed all other automated methods in terms of sensitivity (0.998), specificity (0.998), precision (0.997), and F1 score (0.993). Each method took under 5 minutes to identify and remove duplicates.

Deduplication method	*Duplicate citations removed*	*Citations remaining*
TRUE duplicates (all methods + hand searching)	*1,293*	*2,145*
Human	*1,280*	*2,158*
Endnote (automatic)	*986*	*2,452*
SRA-DM	*1,092*	*2,346*
AsySD	*1,282*	*2,156*

Table 4.6: Duplicate citations identified in the neuroimaging dataset by each method

	TP	TN	FN	FP	Sens	Spec	Prec	F1	Time
Human	1,274	2,139	19	6	0.985	0.997	0.996	0.990	?
Endnote	983	2,142	310	3	0.760	0.999	0.997	0.863	<5 m
SRA-DM	1,050	2,103	243	42	0.812	0.980	0.962	0.880	<5 m
ASySD	1,278	2,141	15	4	0.988	0.998	0.997	0.993	<5 m

Table 4.7: Performance of each deduplication tool in the neuroimaging dataset
TP= true positive, TN = true negative, FN = false negative, FP=false positive, Sens = Sensitivity, Spec = Specificity, Prec = Precision, m = minutes

4.4.4 Cardiac Dataset

This cardiac dataset (N = 8,948) contained 3,510 duplicate citations (39.2% of total; Table 4.8). The human reviewer sensitivity was high, and they captured most duplicates (sensitivity = 0.893; Table 4.9). Seventeen records had been removed in error. Endnote missed a substantial portion of duplicates (sensitivity = 0.749). The SRA-DM identified many false positives (n=275) and missed many duplicates (n=2,361). The ASySD tool outperformed other automated methods in terms of sensitivity (0.998) and F1 score (0.998) and was matched by Endnote in specificity (0.999) and precision (0.999). Deduplication took less than 5 minutes using Endnote or ASySD and just under 30 minutes using the SRA-DM (Table 4.9).

Deduplication method	*Duplicate citations removed*	*Citations remaining*
TRUE duplicates (all methods + hand searching)	3510	5438
Human	3153	5795
Endnote (automatic)	2737	6211
SRA-DM	1424	7524
AsySD	3507	5441

Table 4.8: Duplicate citations identified in the cardiac dataset by each method

	TP	TN	FN	FP	Sens	Spec	Prec	F1	Time
Human	3,136	5,421	374	17	0.893	0.997	0.995	0.941	?
Endnote	2,734	5,435	776	3	0.779	0.999	0.999	0.875	<5m
SRA-DM	1,149	5,163	2,361	275	0.327	0.949	0.807	0.466	<30m
ASySD	3,503	5,434	7	4	0.998	0.999	0.999	0.998	<5m

Table 4.9: Performance of each deduplication tool in the cardiac dataset
TP= true positive, TN = true negative, FN = false negative, FP=false positive, Sens = Sensitivity, Spec = Specificity, Prec = Precision, m = minutes

4.4.5 Depression Dataset

The depression dataset (N=79,880) contained 10,059 duplicate citations (12.6% of total; Table 4.10). The human reviewer sensitivity was very high, and they correctly identified most duplicates. Endnote missed many duplicate citations (sensitivity = 0.75; Table 4.11) but was highly specific (specificity = 0.99), removing only five duplicate citations incorrectly. The SRA-DM was highly sensitive (sensitivity = 0.98) but removed a substantial number of false positive duplicates (n=1,348). Overall, ASySD had a higher sensitivity (0.957), specificity (0.999), precision (0.993) and F1 score (0.974) than other automated tools. Deduplication using Endnote or ASySD took less than an hour, while the SRA-DM took approximately 48 hours to complete the process.

Deduplication method	*Duplicate citations removed*	*Citations remaining*
TRUE duplicates (all methods + hand searching)	10,059	69,821
Human	94,18	70,462
Endnote (automatic)	75,36	72,344
SRA-DM	10,796	69,084
AsySD	96,96	70,184

Table 4.10: Duplicate citations identified in the depression dataset by each method

	TP	TN	FN	FP	Sens	Spec	Prec	F1	Time
Human	9,390	69,793	669	28	0.933	0.999	0.997	0.964	?
Endnote	7,531	69,816	2,528	5	0.749	0.999	0.999	0.856	<30 m
SRA-DM	9,448	68,473	611	1,348	0.939	0.980	0.875	0.906	~48 h
ASySD	9,624	69,749	435	72	0.957	0.999	0.993	0.974	<1 h

Table 4.11: Performance of each deduplication tool in the depression dataset
TP= true positive, TN = true negative, FN = false negative, FP=false positive, Sens = Sensitivity, Spec = Specificity, Prec = Precision, m = minutes

4.4.6 Systematic review of systematic reviews dataset

The SRSR dataset (N=53,001) had 16,838 duplicate citations (31.7% of total; Table 4.12). The human reviewer sensitivity was high (sensitivity = 0.990; Table 4.13), capturing nearly all duplicates and outperforming other methods. Endnote lacked sensitivity (0.760) and removed the fewest citations overall. The SRA-DM identified many false positives (n=1868) and lacked sensitivity (0.709). The ASySD tool outperformed other automated methods in terms of sensitivity (0.982), precision (0.999) and F1 score (0.991) and was matched by Endnote on specificity (0.999), with a low false positive rate. Manual deduplication had taken one team member (Zsanett Bahor) approximately 9 hours to complete using Endnote. Automated deduplication via ASySD and Endnote took less than 1 hour, and the SRA-DM took just under 24 hours.

Deduplication method	*Duplicate citations removed*	*Citations remaining*
TRUE duplicates (all methods + hand searching)	16,838	36,163
Human	16,778	36,223
Endnote (automatic)	12,830	40,171
SRA-DM	13,814	39,187
AsySD	16,564	36,437

Table 4.12: Duplicate citations identified in the depression dataset by each method

	TP	TN	FN	FP	Sens	Spec	Prec	F1	Time
Human	16,668	36,053	170	110	0.990	0.997	0.993	0.992	~9 hours
Endnote	12,794	36,127	4,044	36	0.760	0.999	0.997	0.862	<1 hour
SRA-DM	11,946	34,295	4,892	1,868	0.709	0.948	0.865	0.779	<24 hours
ASySD	16,543	36,142	295	21	0.982	0.999	0.999	0.991	<1 hour

Table 4.13: Performance of each deduplication tool in the SRSRs dataset
Sen TP= true positive, TN = true negative, FN = false negative, FP=false positive, Sens = Sensitivity, Spec = Specificity, Prec = Precision, m = minutes, h = hours

4.4.7 Overall performance

Across all datasets, Endnote's automated deduplication function and ASySD had consistently low false-positive rates and high specificity (Figure 4.2). ASySD correctly identified more duplicate citations than Endnote (and often more than the human reviewer). SRA-DM removed more duplicates than Endnote in some cases, but the false-positive rate of SRA-DM was high. Compared with the gold standard omnibus test (candidate duplicates identified by any approach and confirmed following human scrutiny), AsySD falsely labelled 101 citations as duplicates, and human reviewers had falsely labelled 164 citations as duplicates. This gives specificity, across all 5 datasets of 0.9991 for ASySD and 0.9986 for human reviewers; and sensitivity of 0.9775 for ASySD and 0.9522 for human reviewers.

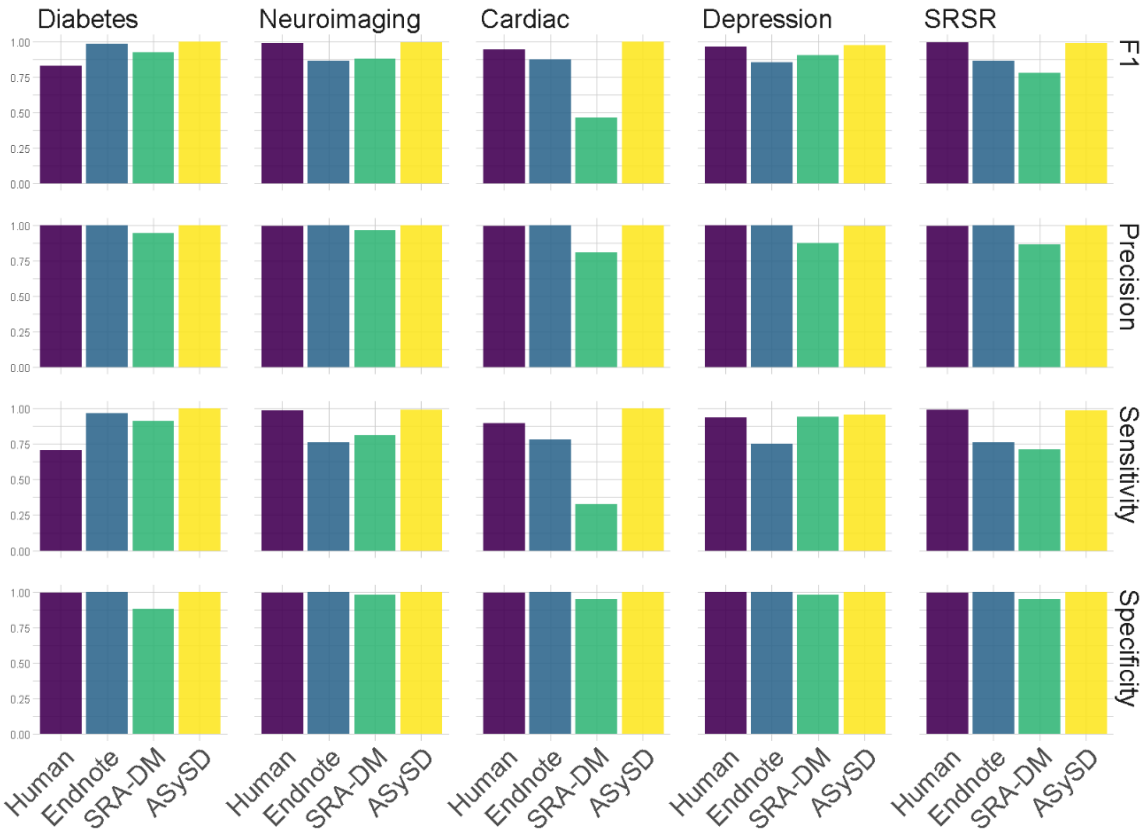


Figure 4.1: Overall performance of different automated deduplication tools and human performance

4.5 Discussion

4.5.1 Human error

I evaluated the performance of different deduplication approaches using datasets from past and existing systematic review projects that were not specifically established to test a deduplication tool. The rationale by which a reviewer removed any given citation is therefore not clear, and there are a number of possible reasons: accidental deletion, removal due to knowledge that article was not relevant, or corrupted files. The process is also likely to be influenced by differences in how reviewers determine what a duplicate is. Often, false positives seemed to be the result of very similar publications (e.g. same title, author, and year) of which

one may be a conference abstract and the other a publication. Information on what would be classed as a "duplicate" was only present in one of the corresponding gold-standard search protocols/publications. For the depression review (Alexandra, Jing, Gregers, & Malcolm, 2016), publications identified in the systematic search which reported the same primary data were considered duplicates, which diverges from my definition.

4.5.2 Dataset Variability

I aimed to test each tool on heterogenous search datasets (i.e. size and number of duplicates) to determine which tool may work best for different types of systematic reviews. Endnote's lack of sensitivity was not immediately apparent on the smallest dataset (Diabetes) but was clearly shown in larger datasets. With the exception of the Diabetes dataset, the sensitivity and specificity of Endnote was fairly consistent across all datasets. The sensitivity and specificity of ASySD was also consistent, indicating that size of dataset and duplicate proportion do not seem to affect performance. SRA-DM varied in performance, with no clear explanatory pattern emerging.

4.5.3 False Positives

While ASySD and Endnote maintained low false positive rates, SRA-DM had a much larger false-positive rate. The SRA-DM was developed on clinical systematic review search datasets, which may differ in key matching criteria or other characteristics. Furthermore, it was previously assessed on 4 relatively small (by preclinical standards) datasets of fewer than 2,000 citations, which may have masked the issue. However, I did not observe trends to suggest that performance was better in smaller datasets compared to larger datasets. I noticed that citations were often removed where they were recorded as having the same DOI. This can occur when a publisher assigns a single DOI to a collection of, for instance, conference abstracts.

In such instances, inspection of the title showed that the works were clearly independent, and not duplicates. The datasets in respect of which this was the biggest problem were also the datasets with the highest proportion of duplicates (Cardiac and Diabetes datasets).

4.5.4 Time taken to remove duplicates

Endnote and ASySD were the fastest methods of deduplication, with all datasets taking under an hour to complete. SRA-DM was extremely slow for larger datasets. However, the interface was user-friendly and if a reviewer is not short of time, the program can run easily in the background without demanding too much processing power.

4.5.5 Limitations OF ASySD

ASySD was developed exclusively using preclinical systematic review datasets. One dataset tested here (Neuroimaging dataset) had both clinical and preclinical studies, however performance has not been evaluated thoroughly on systematic searches within other review areas.

Due to the matching algorithm, the accuracy of ASySD is highly dependent upon the quantity and quality of citation information. All systematic search datasets were from recent reviews, and although each contained older citations, the amount of missing information was relatively low. It is unclear how any of the tools would perform on older searches or citations without page numbers, DOIs, ISBNs, and other useful bibliographic information. In these cases, it is likely that the code may need to be adapted, or that the user would have to supplement with manual deduplication to a greater degree.

Furthermore, ASySD users are likely to have different criteria for determining what counts as a “duplicate”. In future versions of ASySD, I plan to build-in additional

user-defined options to specify whether the algorithm should consider conference abstracts, preprints and journal articles with very similar bibliographic information to be duplicates or not. In time, with machine-readable full-text PDFs, it may also be possible to detect the same data published across multiple publications and flag these as duplicates.

While specificity was comparable to human performance, ASySD did remove some citations incorrectly. For smaller reviews in particular, this risk may not be acceptable, and future versions of ASySD will include the option to manually inspect candidate duplicates.

A key limitation of using ASySD for larger datasets (>50,000 citations) is that the processing requirements outstrip those available in our shinyapps.io subscription. I recommend that users run the application locally in R Studio for this purpose, but understand that this may cause problems for those who are not proficient in R. We are currently exploring alternative approaches which would provide sufficient processing efficiency, such as the development of deduplication software which could be installed locally. I expect that ASySD will be provided on such a platform in the near future. In the meantime, all underlying code for ASySD is openly available and has been formalised into an R package, to ensure it is interoperable and convenient for researchers wishing to integrate ASySD into their own automated evidence synthesis workflows.

4.6 Conclusion

Across five preclinical systematic search datasets of varying size and duplicate proportions, the ASySD tool outperformed the SRA-DM and Endnote in detecting duplicates and had a false-positive rate comparable to human performance. For preclinical systematic reviews, automated duplicate removal using ASySD is a highly sensitive, reliable, and time-saving approach. The ASySD tool is freely available online via a Shiny web application and the code behind the application is open

source. Further research is needed to fully evaluate and disseminate the performance of various deduplication methodologies and prioritise areas for improvement.

CHAPTER 5: SYNAPTIC PLASTICITY AND CONGITION IN TRANSGENIC ALZHEIMER'S MOUSE MODELS

5.1 Chapter introduction

In this chapter, I describe a SR and meta-analysis I performed to summarise the literature using *in vitro* hippocampal slice electrophysiology in transgenic AD models. I attained grant funding (Alzheimer's Research UK) to hire a summer student, Megan McManus to help extract data from included publications. Due to the size of this dataset, (>1400 potentially relevant publications), I prioritised experiments in commonly used APP mouse models of Alzheimer's disease (See Chapter 1 for details on transgenic APP models). I designed the study protocol, developed reviewer training, and managed the project. Most publications in this subset have extracted by both myself and Megan, but we have not yet reconciled our differences so I present only single annotated data here.

5.2 Background

5.2.1 Synaptic plasticity: the neuronal correlate of learning and memory

"Cells that fire together, wire together"
Donald Hebb (1949).

Synaptic plasticity refers to the ability to alter the strength of synaptic connections between neurons. Changes in synaptic efficacy or transmission based on the activity of surrounding neurons is a widely accepted model of learning memory formation (Bear & Malenka, 1994; Bliss & Collingridge, 1993). The underlying theory is that memories are represented in the brain as patterns of activity which are temporally (time) and spatially sensitive. These patterns are thought to be determined by changes in synaptic efficiency i.e. the strength of connection between two neuronal cells. It was discovered that frequent and continuing stimulation by a pre-synaptic

neuron onto a post-synaptic neuron were associated increased efficiency of synaptic transmission (Hebb, 1949).

These ideas are supported by experimental evidence that short trains of high frequency stimulation applied to excitatory pathways in the hippocampus induce an immediate, long lasting increase in synaptic transmission; an effect later conceptualised as long-term potentiation (LTP) (Bliss & Lomo, 1973). Depending on the duration of effect, LTP can be characterised as early (E-LTP) or late (L-LTP), where E-LTP typically lasts <1 hour and L-LTP develops at a later stage and can be maintained for several hours. Importantly, these distinct phases also have different underlying biological processes (Baltaci, Mogulkoc, & Baltaci, 2019). L-LTP relies on protein synthesis and leads to longer-term structural changes. In addition to synaptic strengthening, synapses can also become weakened by infrequent stimulation, though a process known as long-term depression (LTD).

Synaptic transmission is also governed by dynamic, shorter term alterations in synaptic strength. Paired pulse facilitation (PPF) occurs when a pair of stimuli activate the pre-synaptic neuron in quick succession, resulting in an enhancement of the second post synaptic amplitude (Jackman & Regehr, 2017). This form of plasticity is evoked rapidly and then decays within milliseconds or across several minutes (Zucker & Regehr, 2002).

5.2.2 Synaptic failure in AD

Synapse loss is a prominent feature of AD neuropathology. A growing body of evidence suggests that across all AD pathologies, the extent of synapse loss is most correlated to the degree of cognitive impairment (Robinson et al., 2014; Scheff et al., 2006; Terry et al., 1991). Synapse density is highly dependent on synaptic activity (Nägerl, Eberhorn, Cambridge, & Bonhoeffer, 2004). LTP correlates with an enlargement in dendritic spines (Matsuzaki, Honkura, Ellis-Davies, & Kasai, 2004),

while LTD correlates with dendritic shrinkage and synaptic loss (Bastrikova, Gardner, Reece, Jeromin, & Dudek, 2008; Zhou, Homma, & Poo, 2004).

While most novel therapeutics have been focussed on Aβ clearance, some researchers have proposed that AD, in essence, is the result of synaptic failure (Jackson et al., 2019; Selkoe, 2002). In line with other theories of AD pathophysiology, significant *in vitro* evidence suggests that the application of Aβ oligomers induce impairments in LTP (Klyubin et al., 2005; Li et al., 2011; Shankar et al., 2008). Therefore, therapeutic potential may lie in targeting synaptic dysfunction at early stages of the disease (Koffie, Hyman, & Spires-Jones, 2011; Nistico, Pignatelli, Piccinin, Mercuri, & Collingridge, 2012).

5.2.3 Measuring synaptic plasticity

In preclinical research, synaptic plasticity is often assessed by measuring field potentials (from a population of neurons) in *in vitro* hippocampal slices derived from rodent models. LTD is induced by applying low frequency electrical stimulation to synaptic pathways, while LTP can be induced by applying high frequency (e.g. 100 Hz) stimulation. Some induction protocols apply stimulation at frequencies which are aligned to the ordinary firing rate within hippocampal neurons (5Hz), known as theta burst stimulation (TBS). The most common protocol for measuring LTP is to apply excitatory stimulation to axons in hippocampal area CA3 and record their output in CA1, measuring the plasticity of synapses in the Schaffer collateral pathway. LTP changes in other pathways have also been evaluated (see Figure 5.1 for overview of hippocampal pathways). To measure a change in LTP/LTD, baseline field excitatory post-synaptic potentials (fEPSPs) are recorded for a period of time (typically between 1-30 minutes) in the pathway of interest. Once the induction protocol has been applied, experimenters typically report the fEPSPs as a percentage of the baseline potential.

To determine basal synaptic transmission and calculate appropriate baseline stimulation voltages for LTP/LTD experiments, input/output (I/O) relationships are usually plotted. To achieve this, stimulations of increasing voltages are applied, and the resulting output is measured at each step. In most cases, output fEPSP amplitudes are plotted alongside the stimulation intensity or the fiber volley amplitude. The fiber volley is seen as a brief deflection in electrophysiology traces before the larger fEPSP and is an indication of the combined potential of the pre-synaptic neurons.

PPF is typically measured by calculating a ratio of facilitation, by dividing the larger second postsynaptic amplitude by the first. Often the effect is measured across different interstimulus intervals (duration between paired stimuli).

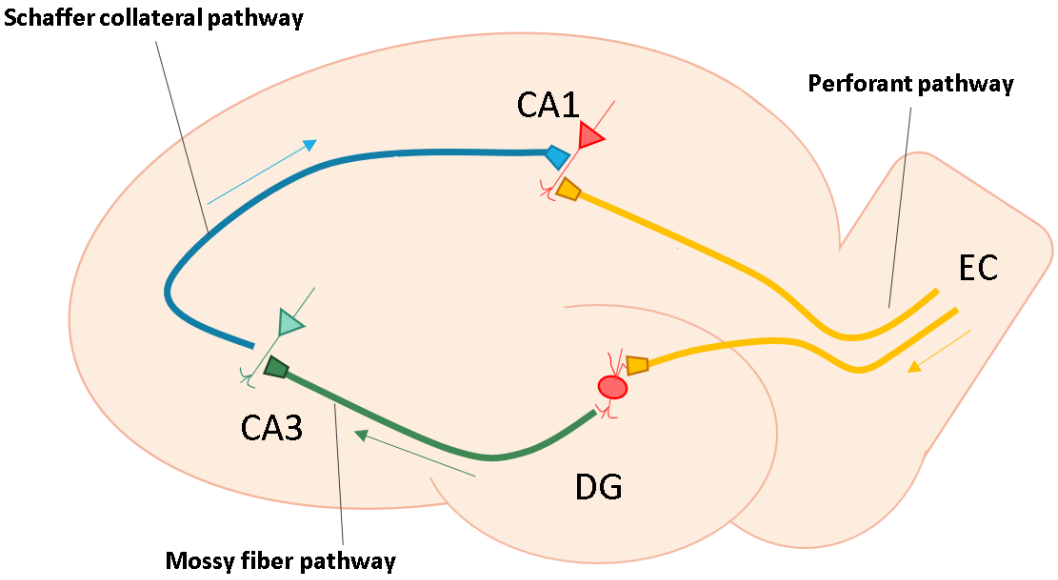


Figure 5.1: Overview of hippocampal excitatory pathways
EC=Entorhinal cortex, DG = dentate gyrus

5.2.4 Reproducibility of *in vitro* electrophysiology experiments

Evidence for synaptic plasticity deficits in hippocampal slices is highly variable and there are many inconsistencies, even within the same AD model (Marchetti & Marie, 2011; Nistico et al., 2012). Conflicting results may be explained by the use of different brain regions, differences in neuronal physiology across AD models, and variations in experimental protocols (Marchetti & Marie, 2011; Sanes & Lichtman, 1999; Tripathy, Burton, Geramita, Gerkin, & Urban, 2015). Indeed, a multitude of study design characteristics have been suggested to affect electrophysiology results, such as the time of day (Chaudhury, Wang, & Colwell, 2005), the temperature of the brain slices during recording (Williams et al., 1993), electrode type (Tripathy et al., 2015), the use of kynureic acid to limit excitotoxicity at brain slicing (Fitzjohn et al., 2001; Hsia et al., 1999), the duration slices are left to recover (Billard, 2010), and the type of anaesthesia used during sacrifice (Zschenderlein, Gebhardt, von Bohlen und Halbach, Kulisch, & Albrecht, 2011)

5.2.5 Synaptic failure and cognitive decline

Given the emerging role of synaptic dysfunction in cognitive impairment in AD, it is highly relevant for drug development pipelines to confirm whether such a link is also present in AD models, however the extent to which synaptic dysfunction contributes to the cognitive behavioural measures of learning and memory in preclinical animal models remains unclear (D'Hooge & De Deyn, 2001).

One of the most common ways to assess learning and memory in rodent models of AD is the Morris water maze (MWM) paradigm (Morris, Garrud, Rawlins, & O'Keefe, 1982; Webster, Bachstetter, Nelson, Schmitt, & Van Eldik, 2014). This test measures spatial learning and memory across two phases. In the acquisition phase, rodents are placed in pool of opaque water with a submerged (hidden) platform which will allow escape. Multiple training sessions are employed to encourage animals to

learn where the platform is and escape from the pool. Longer durations to find the hidden platform indicates an impairment of learning abilities. In the second stage, known as the probe phase, the hidden platform is removed, and animals are placed back into the pool. Longer periods of time spent in the area around where the platform was suggest that the memory is intact, while longer times spent in other segments of the pool indicates that the memory may be impaired.

Although generally thought to be a robust measure (D'Hooge & De Deyn, 2001) the sensitivity of the test can also be influenced by experimental variables, including water temperature, and training protocols during the acquisition phase (Egan et al., 2016; Rubinow, Arseneau, Beverly, & Juraska, 2004)

The current systematic review therefore aims to collect and summarise all preclinical AD studies investigating in-vitro electrophysiology outcomes in combination with behavioural assessments using the Morris Water Maze paradigm, to assess the quality of research and the prevalence of risks of bias in this research field, and to determine which aspects of experimental design are important determinants of the resulting data.

5.3 Methods

Prior to data extraction, I pre-registered a SR protocol on the Open Science Framework (Appendix C5.1) detailing my aims and methodology.

5.3.1 Study identification

Publications with experiments mearing hippocampal synaptic plasticity/transmission *in vitro* were identified within the larger preclinical AD dataset (n = 22,375 publications) described in detail in Chapter 2. I developed a regular expression dictionary (Appendix C5.2) to identify *in vitro* electrophysiology outcomes within the full-text of the publications. To develop this dictionary, I

identified relevant terms from the methods sections of *in vitro* electrophysiology studies included in a previous literature review (Marchetti & Marie, 2011). I used the AutoAnnotation R package (Liao, 2017) to count the frequency of matching terms within each publication. Based on initial testing, publications which reported electrophysiology outcomes had several matches to the regex dictionary. I therefore prioritised studies which had more than three matching terms. I also developed a MWM regex (Appendix C5.3) to identify publications with MWM experiments to pilot the relevant data extraction questions.

5.3.2 Pilot study

I selected a random sample set of 10 papers that had greater than three matches for both the MWM and *in vitro* electrophysiology dictionaries to be part of a pilot study. I uploaded an XML containing the 10 studies to the SyRF platform. Data was extracted by a single reviewer (Myself). This allowed me to determine the feasibility of the data extraction template I had designed, to develop training materials for reviewers (Appendix C5.4), and to identify any potential issues prior to finalising the SR protocol.

5.3.3 Study inclusion

I included studies in transgenic rodent models of AD, with no restrictions on animal species or stage of development, which recorded field potentials to measure synaptic plasticity or transmission in hippocampal slices. I also included studies which investigate the efficacy of treatments on AD models. I defined the control population as a cohort of wild-type animals from the same litter as transgenic animals or an age-matched wild-type of the same background strain which are naïve to any intervention. For treatment efficacy studies, I defined the control population as a transgenic AD model cohort which has not received any treatment

intervention or has received appropriate sham treatment (e.g. injection of vehicle).

Table 5.1 details my inclusion criteria in full.

	Inclusion criteria	Exclusion criteria
Type of 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.
Type of animal model	Transgenic AD animal rodent models 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.
Type of intervention	All primary experiments in transgenic AD rodent models, including studies which test the effect on a treatment given *in vivo*	Studies which only test the effect on a treatment *in vitro*, with no controlled *in vivo* data (e.g. transgenic model + treatment vs transgenic model + vehicle)
Outcome measures	Studies which measure field synaptic transmission (I/O relationships) or plasticity (LTP/LTD/PPF) using *in vitro* electrophysiology in hippocampal brain slices. I will also include behavioural data from Morris water maze experiments.	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	All publication dates	None

Table 5.1: Inclusion criteria for electrophysiology SR

5.3.4 Research questions

In this review, I aimed to investigate the following questions:

- What experimental design variables are reported in hippocampal slice electrophysiology studies and what is the influence of these factors?
- What is the prevalence of risk of bias reporting in the literature and what is the impact of this on effect sizes?
- What is the impact of modelling (transgenic) or treatment interventions on synaptic plasticity and do these effects align with cognitive behavioural tests (MWM outcomes)?

5.3.5 Study characteristics

The characteristics I extracted from included publications are shown in Table 5.2.

Information extracted	Additional details
Study meta-data	
Title	Obtained from Endnote
DOI	Obtained from Endnote
First author	Obtained from Endnote
Corresponding author	Obtained from Endnote
Year	Obtained from Endnote
Journal name	Obtained from Endnote
Country of origin of corresponding author	
Animal husbandry	
The light cycle	Light cycle hours (e.g. 12) and whether this was a reversed cycle (e.g. light in evening)
The number of animals per cage	
Presence of environmental enrichment	E.g. nesting material or toys in cages
Model induction	
The animal species	
The animal background strain	
The transgene/ genetic manipulation	
The sex of the animals	
The source of animals	
The type of model control	E.g. wild-type littermates or same background strain
Outcome information	
The age of the animals at time of outcome assessment	Where reported as a range, we calculated the mean
Electrophysiology protocols	
Time of day animal sacrificed	
Sacrifice method	With or without anaesthesia
Outcome measure type	LTP, LTD, PPF, I/O, MWM
Calcium concentration of recording solution	

Magnesium concentration of recording solution	
Calcium concentration of slicing solution	
Magnesium concentration of slicing solution	
Presence of kynurenic acid in slicing solutions	
Length of time brain slices left to recover prior to stimulation	
Temperature slices left to recover at	
Type of recording chamber	E.g. submersion, interface chamber, or multi-electrode array
The brain pathway recorded	E.g. schaffer collaterals CA1-CA3
Time taken to record baseline activity before LTP	
Presence of blockers used during recording	
Percentage of maximal response used for baseline (e.g. % of input-output)	% maximum input-output
Induction protocols for LTP and LTD recording	Type of stimulation, number of stimulations
Percentage of maximal response used for LTP/LTD induction	% maximum input-output
Morris water maze protocols	
Water temperature of the pool	
How experimenters record activity	Automated or manual assessment
Habituation time to pool	
Training protocol for acquisition phase	Days/trials used to train animals to locate the platform
Timing of acquisition and probe phase	
Treatment information	
Drug dose	
Drug dose units	
Number of times drug was given	
Route of delivery	
The length of drug treatment	
The age of the animals at time of treatment	Where reported as a range, we calculated the mean
Time between the administration of the treatment and outcome measurement	
Details of anaesthesia (if used for treatment)	

Risk of bias and methodological quality
Reporting of random allocation of animals to treatment/control groups
Reporting of blinded assessment of outcome
Reporting of animal/data exclusions
Reporting of a sample size calculation
Reporting of approval by animal welfare committee
Reporting of a potential conflict of interest
Whether a study protocol is available dated before the experiments began

Table 5.2: Information extracted from publications measuring synaptic plasticity in hippocampal slices

5.3.6 Methods for data extraction

I extracted whether the sample size (N) was reported as slices and/or animals. Using slices as an experimental unit for studies where the intervention took place *in vivo* is a form of pseudoreplication (Lazic, Clarke-Williams, & Munafò, 2018) Slices taken from the same animal are highly likely to be dependent on one another.
As pseudoreplication can severely weaken the conclusions of statistical tests (Lazic, 2010), I extracted information from publications which fell into this category but did not include them in the meta-analysis.

I uploaded publications identified using the *in vitro* electrophysiology regex to SyRF for data extraction and extracted study characteristics and numerical outcome data (mean, SD or SEM, and sample size) from included publications. When outcome data were presented in a graphical format, I used Web Plot Digitizer (https://automeris.io/WebPlotDigitizer/) to measure means and errors bars. Data was extracted by myself and/or my summer student, Megan McManus. Any errors detected during data cleaning were corrected in SyRF prior to analysis. For any publications which omit vital information, I recorded the publication information on

a shared spreadsheet, so that authors could be contacted later. This process will be completed by the time the entire review is fully reconciled.

5.3.7 Methods for selecting APP modelling sub-set

To select a subset of data for analysis, I wrote a custom R code to retrieve and format data from SyRF, clean the data, select studies where the model was one of the commonly used APP models discussed in Chapter 1 (Table 1.1). I also summarised the data, generated visualisations, and performed the meta-analysis within R. This code-based method was preferred over other approaches as the results are reproducible and my decision-making process is transparent.

5.3.8 Random effects meta-analysis for electrophysiology outcomes

I analysed each electrophysiology outcome (LTP, LTD, PPF, I/O) measured in each hippocampal pathway separately where there was sufficient data (N>10 publications and N>25 comparisons). To pool LTP or LTD outcomes in the analyses, I had originally intended to calculate an area under the curve by extracting multiple time points over the experiment (as specified in the SR protocol; Appendix C5.1). However, when extracting from figures within publications, datapoints were often unreadable due to overlapping error bars. Due to the time taken to extract large numbers of datapoints for each experiment and the uncertainty in these measures, I decided to instead take a point measurement every 30 minutes where this was clearly visible in the graph. As previous research has identified disparities in E-LTP and L-LTP across the rodent lifespan and differential relationships to cognitive measures (Huang & Kandel, 2006), I aimed to investigate LTP stages separately. The 30 minute and 60 minutes timepoints are most likely a measure of E-LTP, while beyond the 60 minute timepoint could represent a measure of L-LTP.

To pool experiments measuring I/O relationships, I extracted the maximal I/O value in the curve reported for each experiment. Maximal I/O values are typically where the largest difference in basal synaptic strength can be observed between cohorts. For PPF, I chose the most commonly reported interstimulus interval and included all experimental data using that interval.

I calculated a standardised mean difference (SMD) for each individual modelling versus wild-type comparison for each outcome measure. Occasionally, electrophysiology outcomes in the same cohort of animals were measured using different units (e.g. I/O relationships measuring fEPSP / fiber volley amplitude or fEPSP / stimulus strength). In these instances, I grouped together the outcomes and calculated a nested effect size in the analysis according to previously described methods (Vesterinen et al., 2014).

When a single wild-type control group served multiple transgenic modelling groups, I adjusted for in the analysis by dividing the size of the control group by the number of groups it serves. Following methods described previously for preclinical meta-analysis (Vesterinen et al., 2014), I calculated Hedge's G effect sizes to obtain a standardised mean difference (SMD) for each modelling comparison (transgenic group versus control group). Effect sizes were weighted based the standard error of each study, with more precise studies given greater weight in the meta-analysis. I pooled SMD effect sizes for each outcome using a random-effects model with a restricted maximum likelihood (REML) estimate of between study variance to get an overall effect size.

5.3.9 Meta-regression analysis: synaptic plasticity and cognition

I calculated a standardised mean difference (SMD) for each individual modelling versus wild-type comparison for MWM outcomes and split the dataset into acquisition and probe phase outcomes. If animals were measured over multiple

timepoints (e.g. escape latency over multiple training days), I calculated the area under the curve to obtain a single point estimate for that cohort of animals, as described previously (Vesterinen et al., 2014). 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), I calculated a nested effect size for that cohort. Within the largest available electrophysiology outcome dataset, I identified studies where (1) the MWM was reported to have been conducted in the same cohort of animals or (2) it was possible that MWM outcomes had been measured within the same cohort, but this was unclear within the publication. In the case of the latter, I included experiments which had measured MWM outcomes in rodents of a younger age or at the same age as electrophysiology outcomes. I then performed a univariable meta-regression within the largest electrophysiology dataset with MWM effect sizes (SMD) as the predictor variable to understand whether cognitive outcomes explained a significant proportion of the heterogeneity.

5.3.10 Multivariable meta-regression

I conducted a multivariable meta-regression to identify sources of heterogeneity across *in vitro* electrophysiology experiments. Although univariable analysis is the conventional approach in pre-clinical meta-analysis, a multivariable method was preferred as it has been shown to explain a greater proportion of the heterogeneity (Tanriver-Ayder, Faes, van de Casteele, McCann, & Macleod, 2021). To explore the heterogeneity in the dataset, I performed a model building validation exercise to fit a multivariable meta-regression model following previously described methodology (Harrer, Cuijpers, Furukawa, & Ebert, 2021; Tanriver-Ayder et al., 2021). I began with performing univariable meta-regression to determine the proportion of heterogeneity accounted for by each variable. The most significant variable was included in the model, and then other covariates were added sequentially in

combination with the first variable. To determine if another covariate should be added, I recorded the change in Akaike's information criteria (AIC) by running a likelihood ratio test to compare models. Where the new model (with the covariate) lowered the AIC and the likelihood test was significant, this indicates an improvement in the model i.e. it explains a greater proportion of the heterogeneity. For each interaction, I added the covariate which lowered the AIC by the greatest amount and then repeated the process by adding another covariate. The model was considered complete when the AIC showed no further reductions on the addition of new covariates, or any additional variables did not significantly improve model fit.

If a variable is missing, the experimental comparison (transgenic model versus control) will be removed from the dataset when conducting a meta-regression analysis. If a variable was reported in less than 25 experiments, I did not include it as a covariate in the multivariable meta-regression model. Further, if a numerical continuous variable was missing in more than 10% of experiments, I did not include it as a covariate in the multiple meta-regression model as this was deemed to remove too many potentially useful comparisons from the analysis. For categorical variables, if a category applied to less than 10 experiments, it was grouped into an "Other" category or combined, where reasonable, with another category. For categorical variables with less than 10 in one category and only two meaningful categories e.g. "Reported" and "Not reported", that variable was not included in the heterogeneity analysis as there would not be enough statistical power to detect an effect.

5.3.11 Publication Bias

I assessed the extent of publication bias in the dataset using funnel plots, Egger's test and trim and fill analysis.

5.4 Results

5.4.1 Study identification

The electrophysiology regex, with a criterion of three or more matches, identified 1,455 potentially relevant publications within the preclinical AD dataset. A small number of additional duplicates were identified during data extraction or analysis and removed from further evaluation, leaving 1,431 publications in the dataset.

I identified 166 publications describing modelling and intervention experiments in selected APP transgenic models where synaptic plasticity / transmission was evaluated in hippocampal slices. I split the extracted data into two datasets – a modelling dataset comparing APP transgenic animal models groups to an appropriate wild-type control (n=151) and a treatment dataset comparing treated APP transgenic model groups to APP transgenic untreated controls (n=74). For modelling experiments, I included experiments where a vehicle or sham treatment had been given to both wild-type and transgenic animal groups. A PRISMA flow diagram of included studies for the sub-set is shown in Figure 5.2 and a list of all included publications is shown in Appendix C5.5.

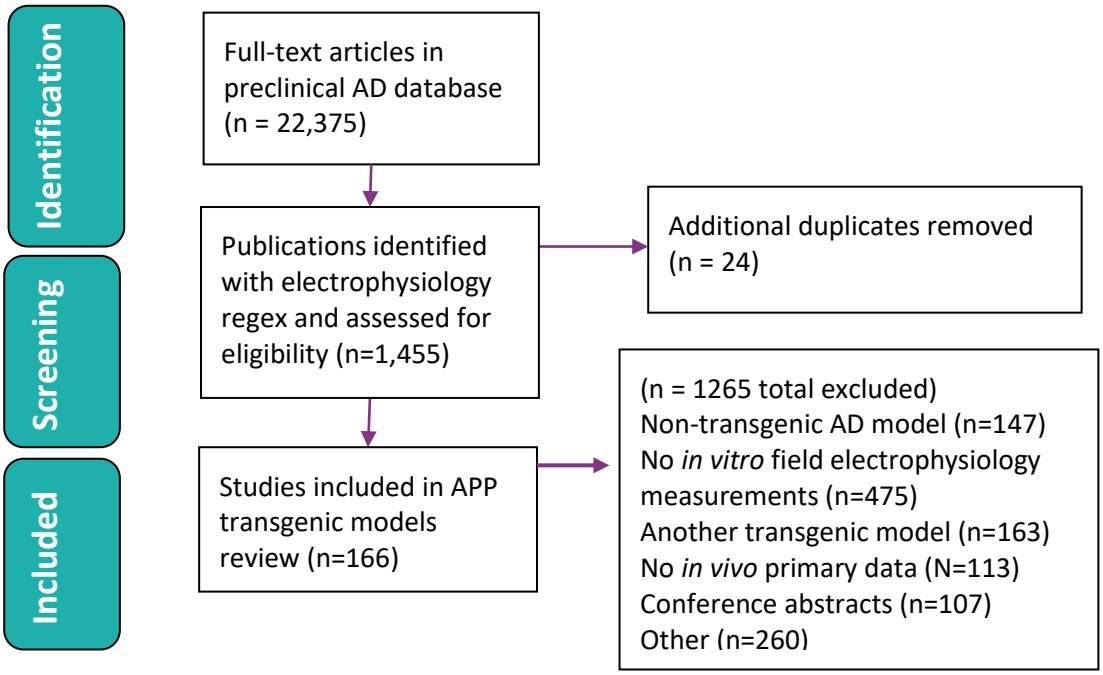


Figure 5.2: PRISMA flow diagram for electrophysiology review

Note: many papers will be excluded for several reasons. The order of exclusions indicates the main reasons, where a paper will be excluded for not being in a non-transgenic model before not recording field potentials in vitro

5.4.2 Publications over time

The publication rate of experiments measuring synaptic plasticity in hippocampal slices from transgenic APP rodent models has increased over time. There were 27 relevant publications in 2017 (Figure 5.3), while there were less than 10 publications a year up until 2010. Since then, there has been an average of 19 new publications per year. The number of papers published in 2018 is inappropriately small due to the timing of the search (January 2018).

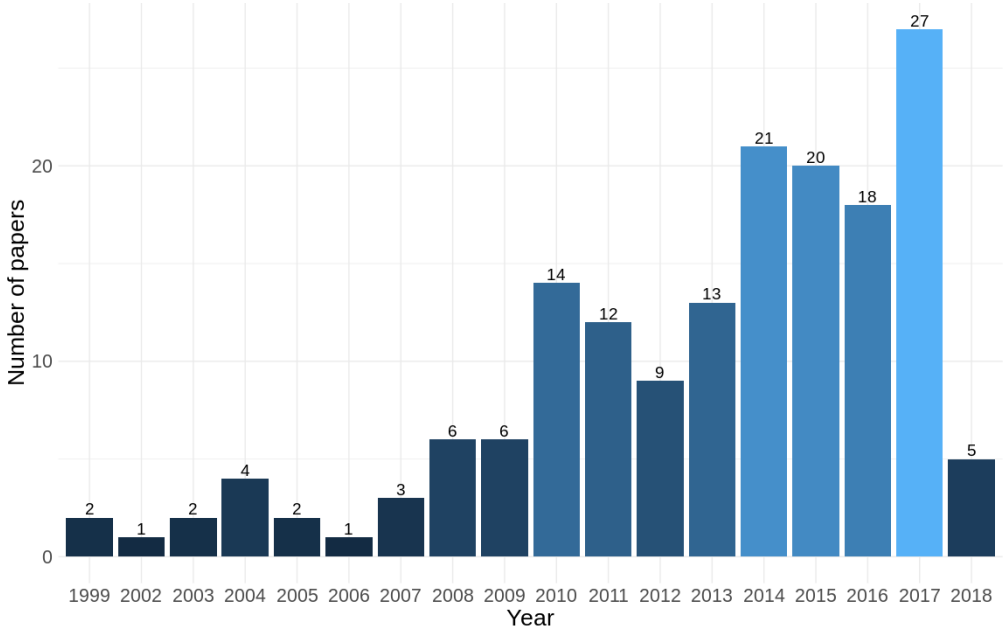


Figure 5.3: Included publications per year in electrophysiology review

5.4.3 Research location

The most common corresponding author location for included publications was the United States (N=79), followed by China (N=22), and Italy (N=11). A map indicating the number of publications from each country is shown in Figure 5.4. This is likely to reflect where research has been indexed within the biomedical databases I

searched. It may not be a true representation of where most research is being conducted.

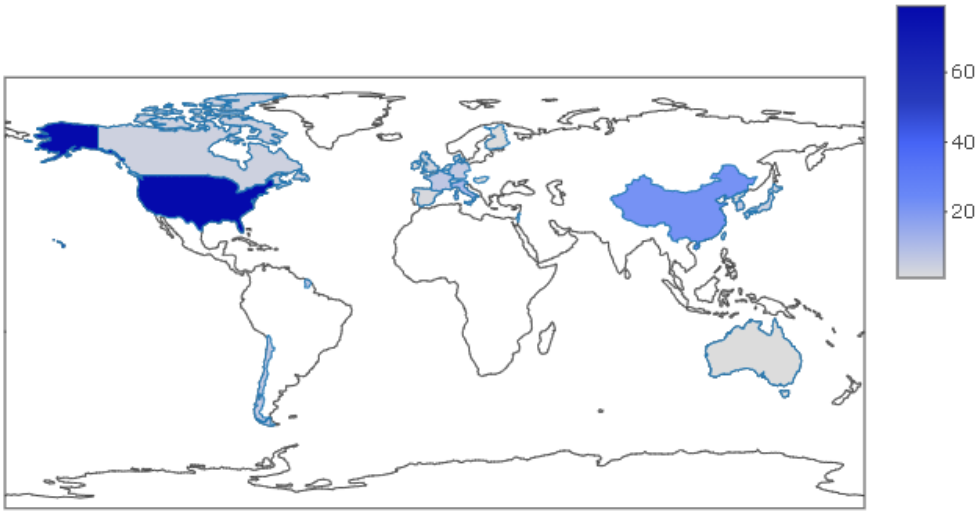


Figure 5.4: Corresponding author location in electrophysiology review
Daker shades of blue indicate a higher frequency of publications

5.4.4 Transgenic model characteristics

A summary of the number of publications with modelling and treatment experiments using each transgenic model is listed in Table 5.3. The APPSwe/PSEN1de9 model was most frequently used in the modelling dataset (n=45) and treatments were also most frequently tested in this model. Overall, for the modelling dataset, there were n=151 publications describing modelling experiments in common transgenic APP models. Of these, n=58 used single APP models, n=73 publications in APP/PS1 models, and n=20 in APP/PS1/MAPT models. In the treatment dataset (n=73 publications), there were n=20 publications in single APP models, n=40 in APP/PS1 and n=13 in APP/PS1/MAPT models. A full list of publications, models used, and sex of the animals is shown in Appendix C5.7. The background strain of transgenic animals was reported in just over half of publications (86/166). The most common strain was the C57BL/6 mouse, reported in 37/86 publications.

Model	Mutations	Modelling dataset (n=151)	Treatment dataset (n=74)
PDAPP	APP	2	1
APP23	APP	3	1
TgCRND8	APP	11	3
J20	APP	14	3
Tg2576	APP	28	12
PS/APP	APP/PS1	1	0
APPSwe/PSEN1(A246E)	APP/PS1	4	0
APPPS1	APP/PS1	7	5
5xFAD	APP/PS1	16	7
APPSwe/PSEN1dE9	APP/PS1	45	28
3xTg-AD	APP/PS1/MAPT	20	14

Table 5.3: Summary of models used in included electrophysiology publications

5.4.5 Interventions

A total of 74 publications in the dataset tested the effect of a treatment intended to improve AD pathology. Only one treatment (Fluoxetine) was reported in more than one publication. I did not to perform a meta-analysis of this dataset due to the low sample size and variation in transgenic models on which these treatments were tested. Treatments for each publication are listed with transgenic models in Appendix C5.6

5.4.6 Outcome measures

There were two different outcome measures reported for LTP, two for LTD, nine different measures of input/output relationships, and four outcomes measuring PPF (Table 5.4). The majority of papers reported the normalised field extracellular field potentials (fEPSP) slope as a percentage of baseline (pre-LTP) values (reported in

137/151 papers in the modelling dataset and 64/74 papers in the treatment dataset).

Outcome(s) measured	Number of papers in modelling dataset	Number of papers in treatment dataset
LTP		
Normalised fEPSP slope (%)	137	64
Normalised fEPSP slope (Proportion)	6	6
LTD		
Normalised fEPSP slope (%)	9	3
Normalised fEPSP slope (Proportion)	0	1
Input/Output		
fEPSP amplitude / stimulation intensity	6	3
fEPSP slope / stimulation intensity	47	10
fEPSP slope / fiber volley amplitude	23	8
Fiber volley amplitude / stimulation intensity	6	4
Maximum I/O slope	4	0
Minimum I/O slope	1	0
Normalised fEPSP / stimulation intensity	1	1
fEPSP amplitude / current	0	1
fEPSP Population spike	1	0
Paired pulse facilitation		
Paired pulse ratio	52	14
Facilitation (%)	1	0

Outcome(s) measured	Number of papers in modelling dataset	Number of papers in treatment dataset
Increase in facilitation (%)	1	0
fEPSP slope (%)	2	0

Table 5.4: Summary of electrophysiology outcomes reported across publications

Overall (see Table 5.5), although 156/166 papers reported LTP measurements, only 81/156 (51.6%) reported a corresponding I/O relationship as a measure of basal synaptic transmission. Eleven papers measured LTD, and only seven reported both LTP and LTD measurements. I did not identify any publications which measured LTD outcomes in the J20, APPPS1, APP23, or PDAPP transgenic models. PPF had been measured in 58 publications and across every commonly used APP transgenic model except APP23. Three publications had measured I/O, LTP, LTD, and PPF within the same model.

The Morris water maze was used to measure cognition in 53/166 included publications, however, it was only clear in seven publications that experiments had used the same cohort of animals as the electrophysiology experiments. Thirty-eight papers were unclear on whether the same cohort of animals had been used or not.

ID	Model	IO	PPF	LTP	LTD	*MWM*
101	Tg2576	Yes	Yes	Yes	No	*No*
102	Tg2576	Yes	No	Yes	No	*No*
103	Tg2576	Yes	Yes	Yes	No	*Unclear if same cohort*
104	Tg2576	Yes	Yes	Yes	No	*No*
105	Tg2576	Yes	Yes	No	No	*No*
106	Tg2576	Yes	Yes	No	No	*No*
107	Tg2576	Yes	No	Yes	No	*No*
108	Tg2576	No	Yes	Yes	No	*No*
109	Tg2576	Yes	No	Yes	No	*No*
110	Tg2576	Yes	Yes	Yes	No	*No*
111	Tg2576	No	Yes	Yes	No	*No*
112	Tg2576	Yes	Yes	Yes	Yes	*No*
113	Tg2576	Yes	Yes	Yes	No	*No*
114	Tg2576	Yes	No	Yes	No	*Unclear if same cohort*
115	Tg2576	Yes	No	Yes	No	*No*
116	Tg2576	No	No	No	Yes	*No*

117	Tg2576	No	No	Yes	No	*No*
118	Tg2576	Yes	No	Yes	No	*No*
119	Tg2576	Yes	No	Yes	No	*No*
120	Tg2576	No	No	Yes	No	*No*
121	Tg2576	No	No	Yes	No	*No*
122	Tg2576	No	No	Yes	No	*No*
123	Tg2576	Yes	No	Yes	No	*Measured in same cohort*
124	Tg2576	No	No	Yes	Yes	*No*
125	Tg2576	Yes	Yes	Yes	No	*No*
126	Tg2576	No	Yes	Yes	No	*No*
127	Tg2576	Yes	Yes	Yes	No	*No*
128	Tg2576	No	No	Yes	No	*No*
129	Tg2576	Yes	Yes	Yes	No	*No*
130	Tg2576	Yes	Yes	Yes	No	*Unclear if same cohort*
131	Tg2576	No	No	Yes	No	*No*
132	Tg2576	No	No	Yes	No	*Unclear if same cohort*
133	J20	Yes	No	Yes	No	*No*
134	J20	Yes	Yes	Yes	No	*No*
135	J20	Yes	Yes	Yes	No	*No*
136	J20	Yes	No	Yes	No	*No*
137	J20	Yes	Yes	Yes	No	*Measured in different group*
138	J20	Yes	Yes	Yes	No	*Measured in different group*
139	J20	Yes	Yes	Yes	No	*Measured in different group*
140	J20	No	No	Yes	No	*No*
141	J20	Yes	No	Yes	No	*Measured in different group*
142	J20	Yes	No	Yes	No	*Measured in different group*
143	J20	Yes	No	Yes	No	*Unclear if same cohort*
144	J20	Yes	No	Yes	No	*No*
145	J20	No	No	Yes	No	*No*
146	J20	No	No	Yes	No	*Unclear if same cohort*
147	5xFAD	Yes	Yes	Yes	No	*No*
148	5xFAD	Yes	No	Yes	No	*No*
149	5xFAD	Yes	No	Yes	Yes	*Unclear if same cohort*
150	5xFAD	Yes	Yes	Yes	No	*No*
151	5xFAD	Yes	No	Yes	No	*No*
152	5xFAD	No	No	Yes	No	*No*
153	5xFAD	Yes	No	Yes	No	*Unclear if same cohort*
154	5xFAD	No	No	Yes	No	*Unclear if same cohort*
155	5xFAD	Yes	Yes	Yes	No	*No*
156	5xFAD	Yes	No	Yes	No	*No*
157	5xFAD	Yes	No	Yes	No	*No*
158	5xFAD	Yes	No	Yes	No	*Unclear if same cohort*
159	5xFAD	Yes	Yes	Yes	No	*No*
160	5xFAD	Yes	Yes	Yes	No	*No*
161	5xFAD	No	No	Yes	No	*No*
162	5xFAD	Yes	No	Yes	No	*Unclear if same cohort*
163	5xFAD	Yes	Yes	Yes	No	*Unclear if same cohort*
164	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*Yes in same cohort*
165	APPswe/PSEN1dE9	Yes	No	Yes	No	*No*

166	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
167	APPswe/PSEN1dE9	Yes	No	Yes	No	*Yes in same cohort*
168	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*Unclear if same cohort*
169	APPswe/PSEN1dE9	No	No	Yes	No	*No*
170	APPswe/PSEN1dE9	No	No	Yes	No	*No*
171	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
172	APPswe/PSEN1dE9	No	No	Yes	No	*No*
173	APPswe/PSEN1dE9	No	No	Yes	No	*No*
174	APPswe/PSEN1dE9	No	No	Yes	No	*No*
175	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*Measured in different group*
176	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
177	APPswe/PSEN1dE9	No	No	Yes	No	*No*
178	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
179	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*Yes in same cohort*
180	APPswe/PSEN1dE9	No	No	Yes	No	*No*
181	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
182	APPswe/PSEN1dE9	No	Yes	Yes	No	*Unclear if same cohort*
183	APPswe/PSEN1dE9	No	No	Yes	No	*No*
184	APPswe/PSEN1dE9	No	No	Yes	No	*No*
185	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
186	APPswe/PSEN1dE9	Yes	No	Yes	No	*No*
187	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
188	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
189	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
190	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*Unclear if same cohort*
191	APPswe/PSEN1dE9	Yes	No	Yes	No	*Measured in different group*
192	APPswe/PSEN1dE9	No	No	Yes	No	*No*
193	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
194	APPswe/PSEN1dE9	Yes	Yes	Yes	Yes	*No*
195	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*No*
196	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*No*
197	APPswe/PSEN1dE9	No	No	Yes	No	*No*
198	APPswe/PSEN1dE9	No	No	Yes	No	*No*
199	APPswe/PSEN1dE9	Yes	Yes	Yes	Yes	*No*
200	APPswe/PSEN1dE9	No	Yes	Yes	No	*No*
201	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*No*
202	APPswe/PSEN1dE9	Yes	No	Yes	No	*No*
203	APPswe/PSEN1dE9	No	No	No	Yes	*No*
204	APPswe/PSEN1dE9	No	Yes	Yes	No	*No*
205	APPswe/PSEN1dE9	Yes	Yes	Yes	No	*No*
206	APPswe/PSEN1dE9	No	No	Yes	No	*No*
207	APPswe/PSEN1dE9	No	No	Yes	No	*No*
208	APPswe/PSEN1dE9	No	No	Yes	Yes	*No*
209	APPswe/PSEN1dE9	No	No	Yes	No	*No*
210	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
211	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
212	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
213	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*
214	APPswe/PSEN1dE9	No	No	Yes	No	*Unclear if same cohort*

215	APPSwe/PSEN1(A246E)	Yes	Yes	Yes	No	*No*
216	APPSwe/PSEN1(A246E)	No	Yes	Yes	No	*No*
217	APPSwe/PSEN1(A246E)	Yes	Yes	No	No	*No*
218	APPSwe/PSEN1(A246E)	Yes	Yes	Yes	No	*No*
219	TgCRND8	Yes	Yes	Yes	No	*No*
220	TgCRND8	Yes	No	No	No	*No*
221	TgCRND8	No	No	Yes	No	*No*
222	TgCRND8	Yes	No	No	No	*No*
223	TgCRND8	No	No	Yes	No	*No*
224	TgCRND8	No	No	No	Yes	*No*
225	TgCRND8	No	No	Yes	No	*No*
226	TgCRND8	No	Yes	Yes	No	*No*
227	TgCRND8	Yes	No	Yes	No	*No*
228	TgCRND8	Yes	Yes	Yes	No	*Unclear if same cohort*
229	TgCRND8	Yes	Yes	Yes	No	*No*
230	TgCRND8	Yes	No	Yes	No	*No*
231	3xTg-AD	Yes	Yes	Yes	No	*No*
232	3xTg-AD	Yes	No	Yes	No	*No*
233	3xTg-AD	No	No	No	Yes	*No*
234	3xTg-AD	No	No	Yes	No	*No*
235	3xTg-AD	No	No	Yes	No	*No*
236	3xTg-AD	No	No	Yes	No	*No*
237	3xTg-AD	Yes	Yes	Yes	No	*No*
238	3xTg-AD	Yes	No	Yes	No	*No*
239	3xTg-AD	Yes	Yes	Yes	No	*No*
240	3xTg-AD	No	No	Yes	No	*No*
241	3xTg-AD	No	No	Yes	No	*No*
242	3xTg-AD	Yes	Yes	Yes	No	*Unclear if same cohort*
243	3xTg-AD	No	No	Yes	Yes	*No*
244	3xTg-AD	Yes	Yes	Yes	No	*Unclear if same cohort*
245	3xTg-AD	No	No	Yes	No	*Yes in same cohort*
246	3xTg-AD	No	No	Yes	No	*Unclear if same cohort*
247	3xTg-AD	No	No	Yes	No	*No*
248	3xTg-AD	No	No	Yes	No	*Unclear if same cohort*
249	3xTg-AD	No	No	Yes	No	*Unclear if same cohort*
250	3xTg-AD	No	No	Yes	No	*Unclear if same cohort*
251	3xTg-AD	No	Yes	Yes	No	
252	3xTg-AD	Yes	Yes	Yes	No	*No*
253	PS/APP	Yes	Yes	Yes	No	*No*
254	APPPS1	No	No	Yes	No	*Yes in same cohort*
255	APPPS1	Yes	No	Yes	No	*No*
256	APPPS1	Yes	No	Yes	No	*Unclear if same cohort*
257	APPPS1	No	No	Yes	No	*No*
258	APPPS1	Yes	No	Yes	No	*No*
259	APPPS1	No	No	Yes	No	*No*
260	APPPS1	No	No	Yes	No	*No*
261	APPPS1	No	Yes	No	No	*No*
262	APP23	Yes	No	Yes	No	*No*
263	APP23	No	No	Yes	No	*No*

264	APP23	Yes	No	Yes	No	*No*
265	PDAPP	Yes	Yes	Yes	No	*No*
266	PDAPP	No	No	Yes	No	*Yes in same cohort*

Table 5.5: Summary of outcome measures used across in vitro electrophysiology dataset

5.4.7 Age of animals

The median age of mice given a treatment intervention was 4.7 months (interquartile range: 3.0 – 6.88 months). Electrophysiology outcomes were measured at a median age of 6.0 months (3.87 – 9.37 months interquartile range). As seen in Figure 5.5 below, electrophysiology outcomes were often measured early in the mouse lifespan, and in some cases prior to amyloid pathology developing. In the Tg2576 model in particular, most experiments occur prior to 12 months of age and before amyloid pathology has been established. Treatment interventions typically occur prior to amyloid pathology onset.

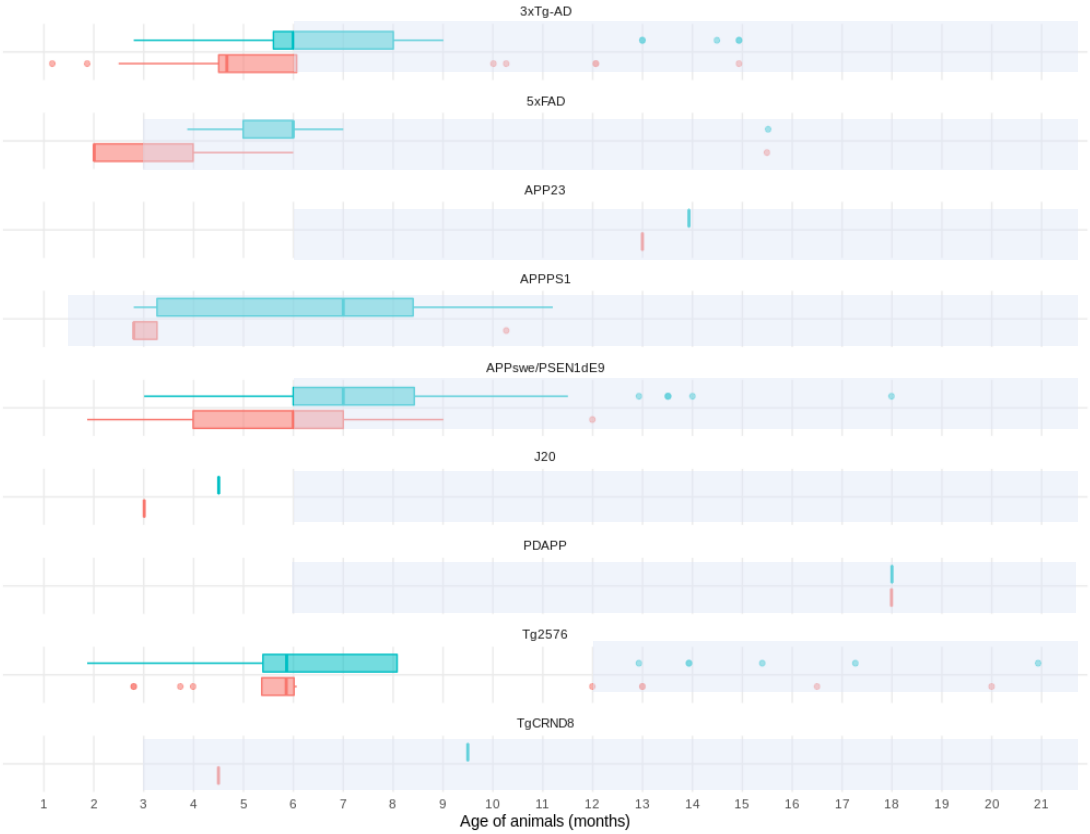


Figure 5.5: Age distribution of mice in electrophysiology experiments

The distribution of age at which electrophysiological outcomes were measured (blue) and age at which treatment was administered (red) are shown for each APP transgenic model. The lilac shaded area indicates when amyloid pathology is present (source for age of amyloid pathology: Alzforum website).

5.4.8 Sex of animals

The sex of the animals used was not reported for at least one comparison in 26.5% of publications (44/166) and not reported at all in 23.5% (39/166) of publications. Thirty-seven publications reported using exclusively mixed sex groups (22.3% of publications). Only two publications analysed male and female mice cohorts separately. In total, 72 publications (43.4%) measured at least one outcome on male mice, compared to just 18 publications measuring at least one outcome in female mice. The sex of animals reported in each publication is shown in Appendix C5.7.

5.4.9 Study quality and risk of bias

Reporting of study quality and measures to reduce the risk of bias were moderate to poor (Table 5.6). Only four studies (2.4%) reported a sample size calculation to ensure their experiments were adequately powered. Further, just under a third of studies (31.3%) reported that outcome assessors were blinded to experimental groups. Most studies reported a conflict of interest statement (62.3%) and most had reported welfare committee approval (69.9%). It is not possible to assign animals randomly to the transgenic model (this is done naturally). Therefore, randomisation and allocation concealment were only applicable for publications reporting a treatment comparison. Of these, only 18.9% reported randomising animals to groups and 10.8% reported that group allocation was concealed. RoB reporting per publication is shown in Appendix C5.9.

	N reporting	N total	Percentage reporting
Allocation concealment	8	74	10.8%
Blinding	52	166	31.3%
Conflict of interest statement	103	166	62.3%
Exclusion criteria	16	166	9.6%
Randomisation	14	74	18.9%
Sample size calculation	4	166	2.4%
Welfare committee approval	116	166	69.9%

Table 5.6: Reporting of study quality items and measures to reduce the risk of bias in electrophysiology publications

5.4.10 Animal husbandry

Reporting of animal husbandry measures was moderate to poor across included publications (Table 5.7). No study reported details of any environmental enrichment. Relatively few publications (28/166) reported the number of animals housed together in a cage. A larger proportion of publications reported the light cycle of the laboratory animal facilities.

	N reporting	N total	Percentage reporting
Environmental enrichment	0	166	0 %
Light cycle	75	166	45.2 %
Number of animals per cage	28	166	16.9 %

Table 5.7: Reporting of animal husbandry details in electrophysiology publications

5.4.11 Electrophysiology slicing protocols

Key details of the slicing procedures for electrophysiology experiments were often unclear or omitted from publications. Reporting of slicing protocols are summarised in Table 5.8 and shown by publication in Appendix C5.8. Time of day that animals were sacrificed was only reported in four publications. Whether animals were anaesthetised prior to sacrifice was unclear in the majority of experiments, with 99/166 not reporting these details. Of those which did report anaesthesia use, most reported the anaesthetic agent used (58/65). Three papers reported sacrifice by cervical dislocation but did not indicate whether or not this involved anaesthesia. The time slices were left to recover at was reported in 73.5% of papers. For papers reporting this, the median time slices were left to recover was 60 minutes (IQR: 60-102.5). The temperature slices were left to recover at was reported in 76/166 papers (45.8%). The median temperature reported was 30°C (IQR: 29-32°C). Magnesium and calcium concentrations of the slicing solutions were missing from 49.7% and 50.0% of publications respectively. For publications reporting magnesium concentrations, the median was 2.0mM (IQR:1.3-7.0) and for calcium concentrations, the median was 1.4mM (IQR: 0.5-2.4). Histograms showing the spread of numerical data reported are shown in Figures 5.6 – 5.9.

		Number of papers
Time of sacrifice	Morning	1
	Daytime	3
	Not reported	162
Anaesthetised prior to sacrifice	Yes	65
	Unclear – cervical dislocation	3
	No details reported	98
Kynurenic acid used in slicing solution	Yes	3
	Not reported	163
Time slices left to recover	Reported	122
	Not reported	44
Temperature slices left to recover at	Reported	76
	Room temperature	37
	Ambient temperature	1
	Not reported	52
Magnesium concentration in slicing solution	Reported	84
	Not used in solution	1
	Not reported	82
Calcium concentration in slicing solution	Reported	71
	Not used in solution	12
	Not reported	83

Table 5.8: Reporting of slicing procedures in included publications

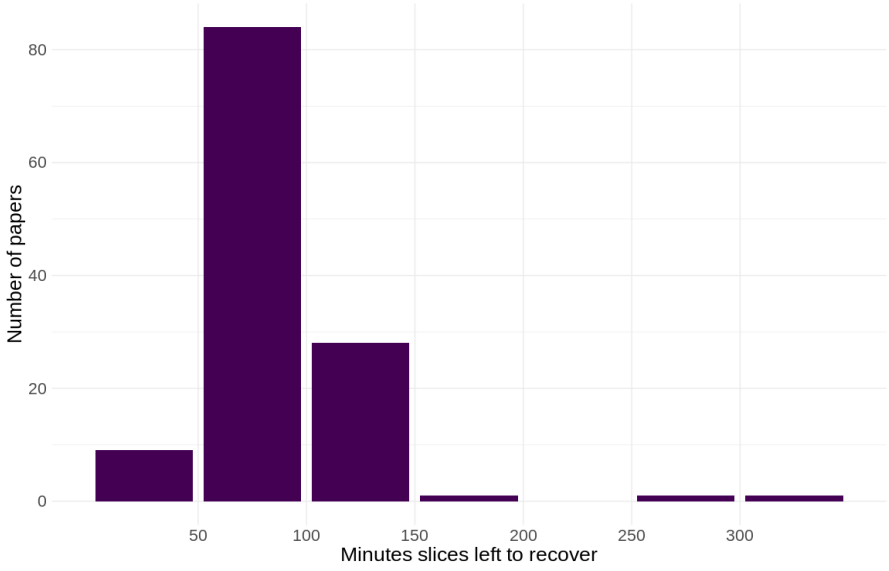


Figure 5.6: Distribution of duration (minutes) slices are left to recover

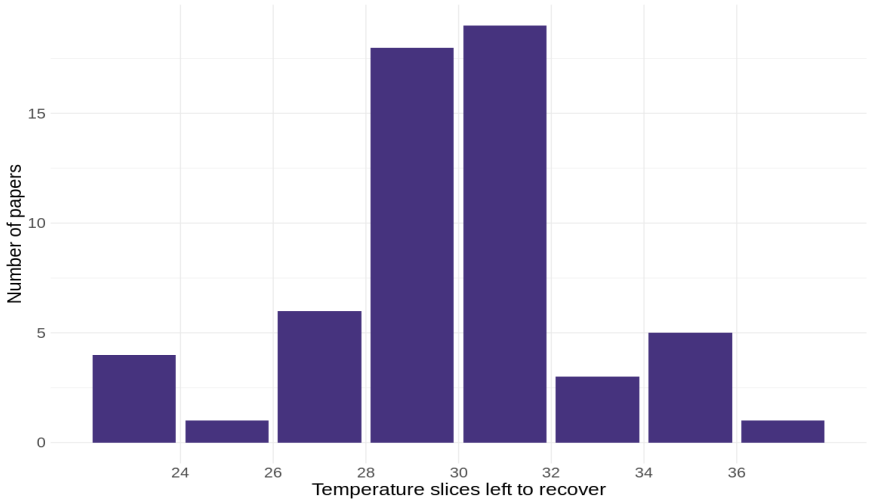


Figure 5.7: Distribution of temperatures(°C) at which slices left to recover

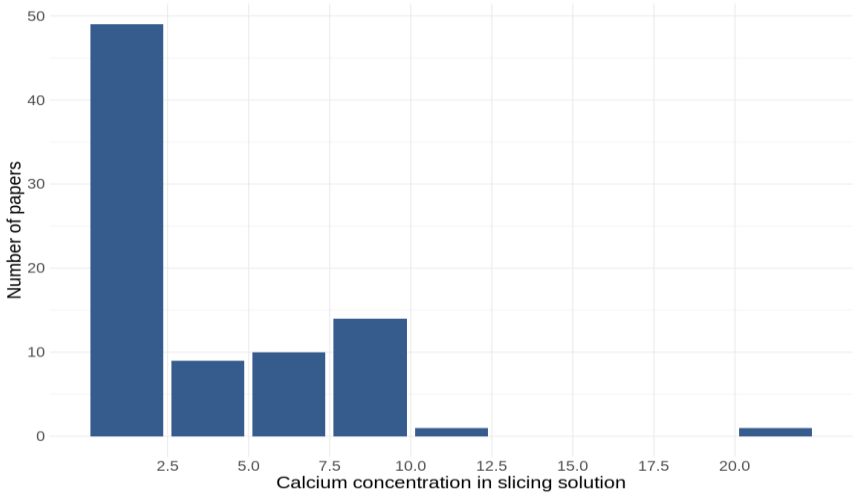


Figure 5.8: Distribution of calcium concentrations (mM) used in slicing solutions

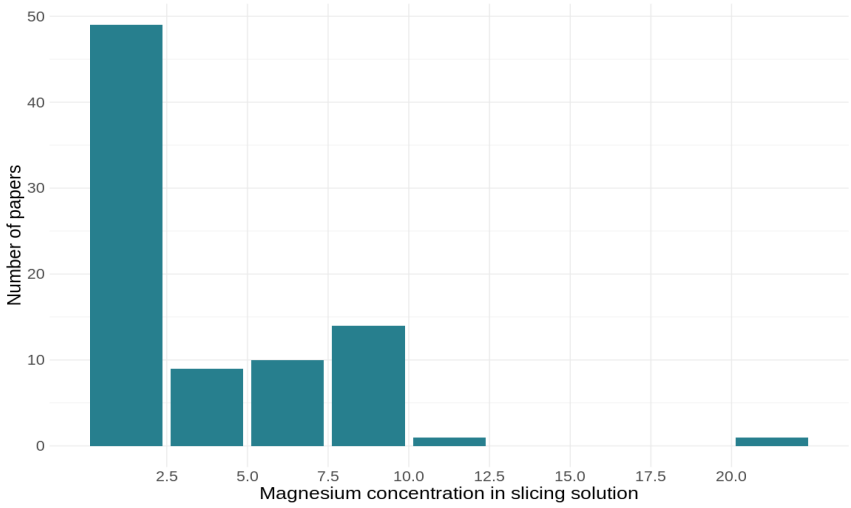


Figure 5.9: Distribution of magnesium concentrations (mM) used in slicing solutions

5.4.12 Electrophysiology recording protocols

The majority (over 90.0%) of papers measured electrophysiological outcomes in the CA1 region of the hippocampus via the Schaffer Collateral pathway. A smaller number of papers (19/166) had experiments measuring electrophysiological outcomes in the Dentate Gyrus via the Perforant pathway. Two papers measured outcomes in the Dentate Gyrus via the Mossy fibre pathway. A substantial proportion of papers (80/166) did not report the way slices were recorded from within the recording chamber i.e. submerged in solution (submersion chamber), interface recording set-up, or multi-electrode array (MEA). Of those reporting the recording set-up, 36/85 used a submersion chamber, and 37/85 used an interface chamber. The remaining publications (12/85) used a MEA recording set-up. Magnesium and calcium concentrations of the recording solution were reported in just over half of papers (85/166). Two papers did not use calcium in their recording solutions. The median magnesium concentration was 1.5mM (IQR: 1.2-2) and median calcium concentration was 2.0mM (IQR:2.0-2.5). The distribution of recording concentrations reported are shown in Figures 5.10 and 5.11.

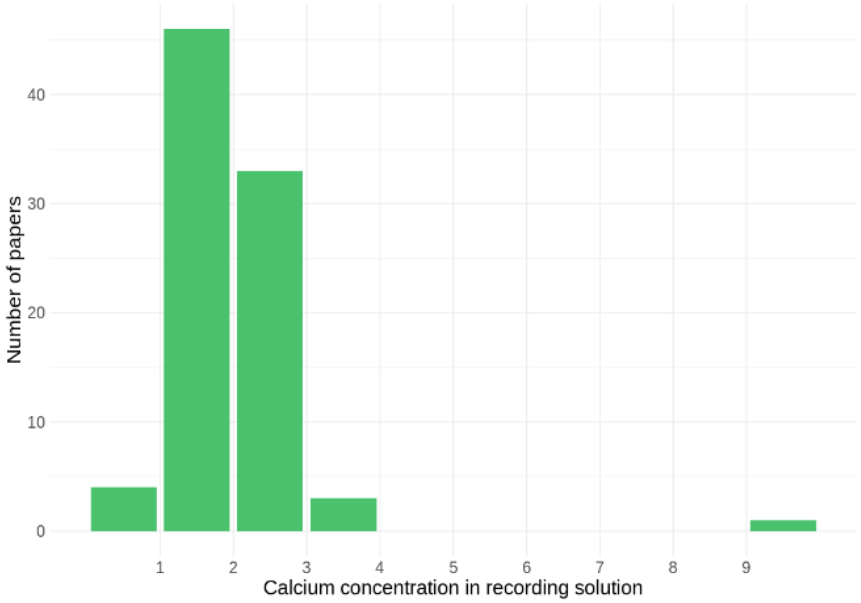


Figure 5.11: Distribution of calcium concentrations (mM) used in recording solutions

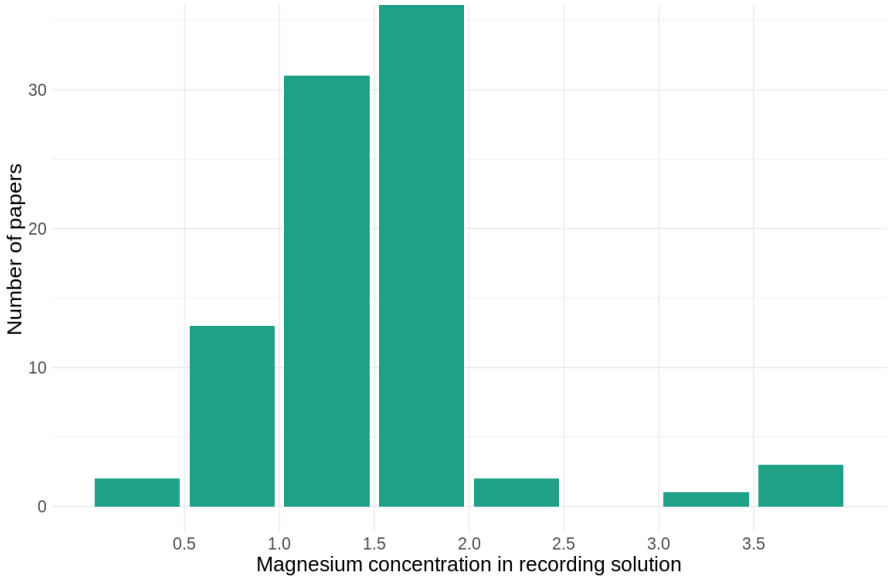


Figure 5.10: Distribution of magnesium concentrations (mM) used in recording solutions

5.4.13 LTP Protocols

Across a total of 253 experiments (across 156 publications) which induced LTP in hippocampal slices, 44.7% (113/253) used theta burst stimulation protocols, 49% used high frequency stimulation protocols, and one publication used an alternative pairing stimuli protocol. Fifteen experiments did not report the stimulation type used to induce LTP. Most experiments measured synaptic plasticity in the Schaffer collateral pathway (226/253). Few experiments (22/253) reported the use of compounds to block specific types of neurotransmissions (see Table 5.9). The median percentage of maximal I/O used as the baseline stimulation intensity reported was 40% (IQR: 30-45) and the median percentage of maximal I/O used for LTP induction was also 40 (IQR: 33 – 50). The time baseline fEPSPs were measured was reported in most experiments (231/253) and was either present within the main text or clear from figures. The median time measured was 20 minutes (IQR: 15-20).

		Number of experiments
Pathway measured	Mossy fibre pathway	2
	Schaffer collaterals	226
	Perforant pathway	23
	Not reported	2
Type of stimulation used	HFS	124
	TBS	113
	Pairing stimuli (2 Hz, 2 min)	1
	Not reported	15
Blockers used	Bicuculline	5
	Bicuculline + CGP-55845 + D-AP5	2
	Picrotoxin	14
	SR95531	1
	None reported	231
Time baseline measured	Reported	231
	Unclear	22
% IO used for baseline stimulation	Reported	163
	Not reported	90
% IO used for LTP stimulation	Reported	66
	Not reported	187
Number of stimulations	Reported	225
	Not reported	28

Table 5.9: Reporting of LTP experimental protocols

Bicuculline, SR95531, and Picrotoxin are *GABA-A receptor antagonists. D-AP5 is an NMDA receptor antagonist. CGP-55845 is a GABA-A and GABA-B receptor antagonist.*

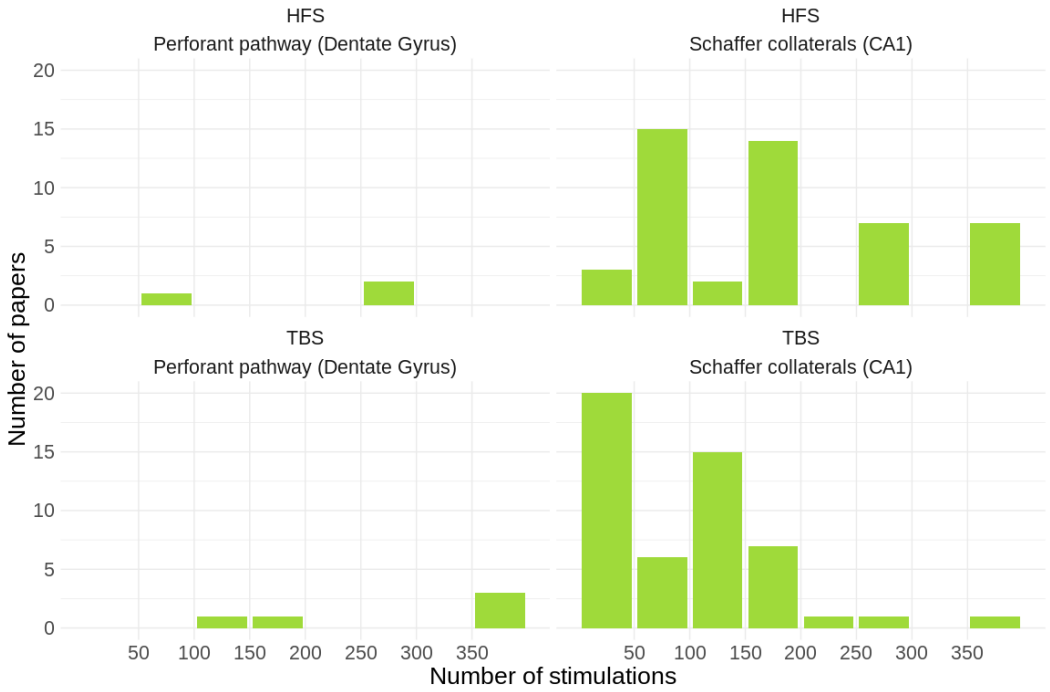


Figure 5.12: Distribution of stimulation protocols for LTP experiments

5.4.14 LTD protocols

Across 11 experiments (in 11 publications) measuring LTD, 10/11 measured synaptic plasticity in the Schaffer collateral pathway. One publication measured LTD in the Perforant pathway. Most applied low frequency stimulation protocols (7/11) using 900 stimulations. Two experiments induced LTD using -3,5-Dihydroxyphenylglycine and one experiment used a strong low-frequency stimulus protocol (2700 stimulations). One experiment used picrotoxin as a blocker of GABA mediated neurotransmission.

5.4.15 MWM protocols

There was variation in MWM protocols across the 53 publications reporting MWM outcomes. Eight reported at least one visible platform trial prior to or after the hidden platform trials. The number of training days for the acquisition phase was reported in every study. The median number of training days was 6 (IQR: 5 – 6).

Most publications reported the number of training trials per day (50/53), with a median of 4 trials per day (IQR: 4-4). By multiplying the number of training days and training trials, I calculated a composite measure of total trials. The distribution of reported training trials used for acquisition in the MWM is shown in Figure 5.13. Twenty-one publications (39.6%) did not report the temperature of the pool. For publications reporting the temperature, the median was 23°C (IQR: 21-24, see Figure 5.14). Most publications reported using various video-tracking software to analyse movement in MWM, while seven publications did not report any details about how this was measured. Nine publications reported habituation trials to the MWM pool. MWM parameters reported per publication is shown in Appendix C5.9.

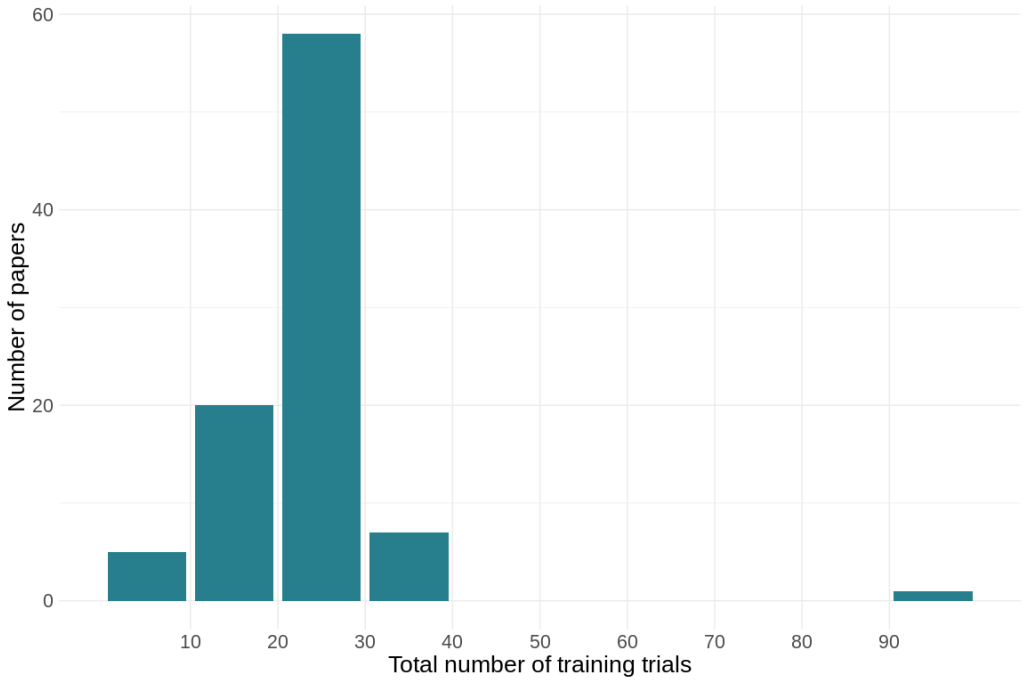


Figure 5.13: Distribution of training trials in acquisition phase of MWM

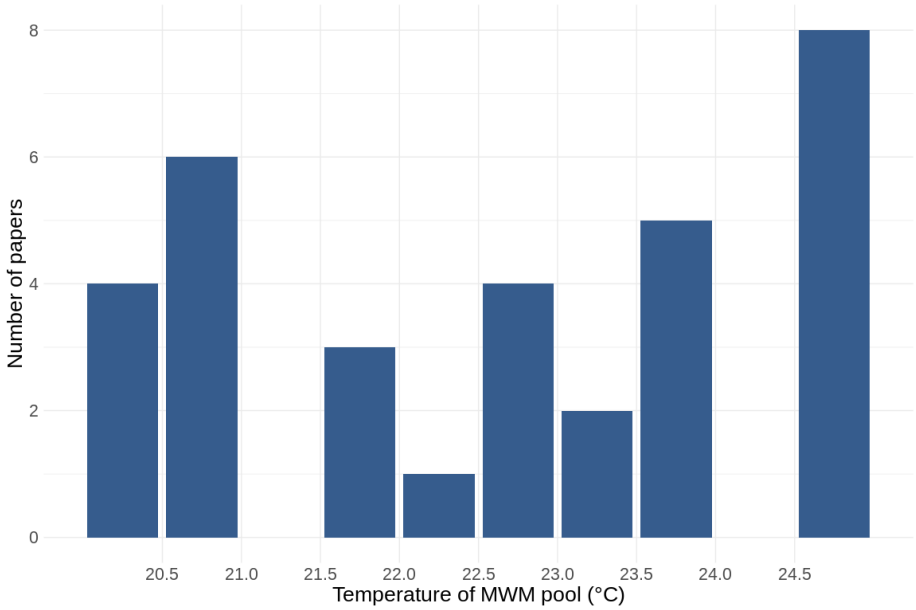


Figure 5.14: Distribution of MWM pool temperatures

5.4.16 Meta-analysis of transgenic modelling interventions

From a total of 151 publications with modelling comparisons, only 106 had useable comparisons for the meta-analysis. Although most studies reported the number of hippocampal slices, 80/151 publications with modelling experiments had at least one comparison which did not report the number of animals used or were unclear about whether the sample size represented slices or animals. The final number of papers and experimental comparisons for each outcome of interest are shown in Table 5.10. Outcomes deemed suitable for analysis (with at least 25 comparisons) are highlighted in lilac.

Outcome	Brain pathway recorded	N comparisons	N publications
Input/Output	Mossy fiber pathway (Dentate Gyrus)	1	1
Input/Output	Perforant pathway (Dentate Gyrus)	10	9
Input/Output	Schaffer collaterals (CA1)	63	43
LTD (30 minutes)	Schaffer collaterals (CA1)	6	3
LTP (30 minutes)	Mossy fiber pathway (Dentate Gyrus)	2	2

LTP (30 minutes)	Perforant pathway (Dentate Gyrus)	11	10
LTP (30 minutes)	Schaffer collaterals (CA1)	92	67
LTP (60 minutes)	Mossy fiber pathway (Dentate Gyrus)	1	1
LTP (60 minutes)	Perforant pathway (Dentate Gyrus)	6	6
LTP (60 minutes)	Schaffer collaterals (CA1)	72	57
LTP (90 minutes)	Perforant pathway (Dentate Gyrus)	1	1
LTP (90 minutes)	Schaffer collaterals (CA1)	13	10
PPF (ISI 50)	Mossy fiber pathway (Dentate Gyrus)	1	1
PPF (ISI 50)	Perforant pathway (Dentate Gyrus)	4	4
PPF (ISI 50)	Schaffer collaterals (CA1)	24	19

Table 5.10 Number of publications and experimental comparisons for each outcome measure

1. ***Long-term potentiation in CA1 at 30 minutes***

In modelling experiments measuring early LTP at a post-stimulation duration of 30 minutes, the pooled effect of transgenic modelling interventions was -0.73 (95% CI –0.95 to -0.51, n=92 comparisons), meaning that transgenic animals had reduced LTP versus wild-type controls. Heterogeneity was high (I^2 = 81.9%, Q=502.40, Tau^2=0.82). A forest plot of comparisons is shown in Figure 5.15 and a forest plot grouped by transgenic model is shown in Figure 5.16

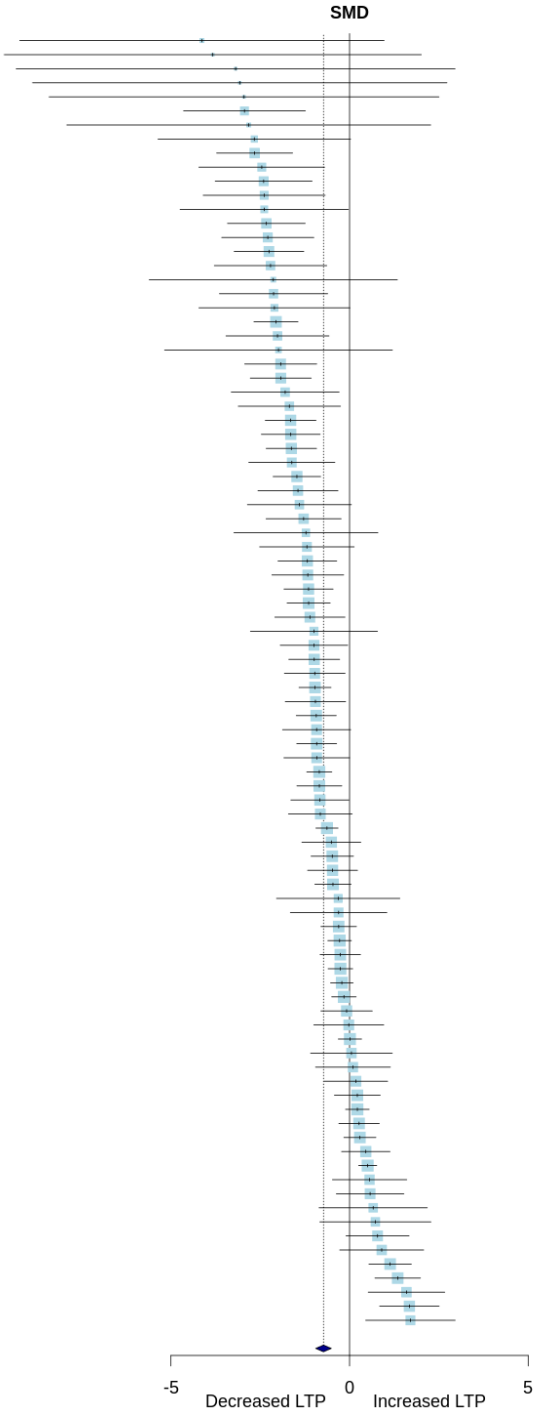


Figure 5.15: Forest plot of random effects meta-analysis of LTP after 30 minutes in modelling experiments.

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. Larger version with more detail available at: https://osf.io/eazxt/

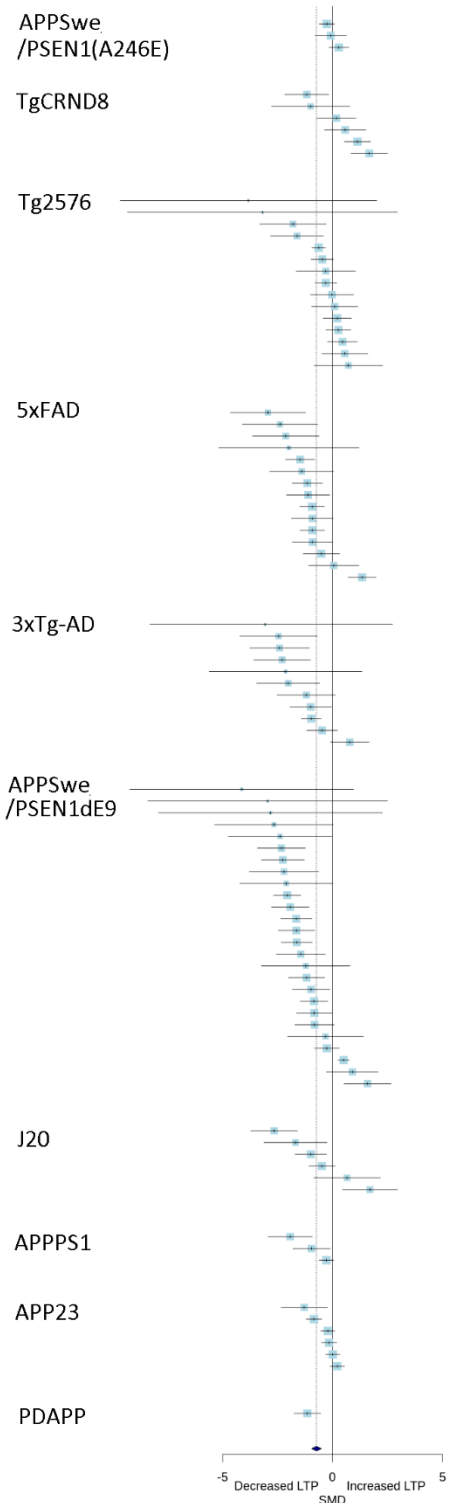


Figure 5.16: Forest plot of random effects meta-analysis of LTP after 30 minutes in modelling experiments, divided by model

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.

In the univariable analysis, several variables each explained a significant proportion of the heterogeneity (Tables 5.11-5.18). Transgenic model explained the highest proportion of heterogeneity (13.81%). Experiments in 3xTg-AD models had the lowest SMD estimates, indicating that 3xTg-AD models have greater reductions in synaptic plasticity (measured by LTP) versus wild type controls. In contrast, experiments in Tg2576 show the least reduction in SMD estimates, which may suggest that synaptic plasticity is less impaired in this model. A similar underlying biological pattern was observed for transgenic model category, where experiments in APP models showed modest reductions in LTP, experiments in APP/PS1 models had larger reductions, and experiments in APP/PS1/MAPT models (e.g. 3xTg-AD) had the largest reduction in LTP.

Reporting of blinded outcome assessment was associated with increased estimates of LTP in transgenic animals (i.e. reduced evidence of LTP impairment). Where conflict of interest statements or compliance with an animal welfare committee was reported, this was associated with a larger reduction in LTP.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
3xTg-AD	-1.23	-1.89 to -0.58	0.0061	0.70	84.39	13.81
Model5xFAD	-0.98	-1.82 to 0.15				
APPswe/PSEN1dE9	-1.14	-1.91 to -0.36				
Other	-0.31	-1.07 to 0.44				
Tg2576	-0.23	-1.07 to 0.61				

Table 5.11: Univariable meta-regression of LTP outcomes at 30 minutes with transgenic model
N=92 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APP	-0.25	-0.59 to 0.08	0.021	0.71	84.75	13.23
APP/PS1	-0.98	-1.43 to -0.53				
APP/PS1/MAPT	-1.23	-1.97 to -0.48				

Table 5.12: Univariable meta-regression of LTP outcomes at 30 minutes with transgenic model category
N=92 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Blinding: Not reported	-0.93	-1.21 to -0.65	0.025	0.77	85.91	5.36
Blinding: Reported	-0.42	-0.86 to 0.03				

Table 5.13: Univariable meta-regression of LTP outcomes at 30 minutes with reporting of blinded outcome assessment
N=92 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
COI statement: Not reported	-0.4	-0.7 to -0.1	0.003	0.72	84.97	12.28
COI statement: Reported	-1.06	-1.48 to -0.63				

Table 5.14: Univariable meta-regression of LTP outcomes at 30 minutes with reporting of conflict of interest statement
N=92 comparisons, CI =confidence intervals, COI= Conflict of interest

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Animal welfare committee approval: Not reported	-0.37	-0.73 to -0.02	0.0145	0.75	85.43	8.87
Animal welfare committee approval: Reported	-0.93	-1.37 to -0.47				

Table 5.15: Univariable meta-regression of LTP outcomes at 30 minutes with animal welfare committee approval
N=92 comparisons, CI =confidence intervals.

The magnesium concentration of recording solution emerged as a significant predictor in the univariable analysis (R^2 = 10.47, p = 0.0417) but was only reported in 54/92 comparisons and could not be evaluated in the multiple meta-regression model due to missing data. As shown in a bubble plot (Figure 5.17), an increase in magnesium concentration by 1mM increased LTP SMD estimates by 0.59. The percentage of maximal I/O used for baseline stimulation (recording prior to LTP induction) also explained a significant proportion of the heterogeneity (R^2 = 13.80%, p=0.01), but was reported in only 64/92 comparisons. A percentage I/O increase of 1% reduced LTP SMD estimates by 0.03. (Table 5.16, Figure 5.18). For the main analysis, I retained the extreme observation where an experiment had used 80% of

maximal I/O for baseline stimulation. However, to assess the potential impact of this, I performed a post-hoc leave-one-out univariable analysis omitting that datapoint. Percentage of maximal I/O was found to explain a slightly larger proportion of the heterogeneity (R^2=18.19%) in this analysis. A percentage I/O increase of 1% reduced LTP SMD estimates by 0.04.

Further, the number of stimulations used to induce LTP explained a significant proportion of the heterogeneity (Table 5.17). A higher number of stimulations was associated with a lower estimate of LTP as shown in Figure 5.19.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Intercept	-1.65	-2.63 to -0.67	0.0417	0.55	84.47	10.47
Magnesium concentration of recording solution	0.59	002 to 1.15				

Table 5.16: Univariable meta-regression of LTP outcomes at 30 minutes with magnesium concentration of recording solution
N=54 comparisons, CI =confidence intervals.

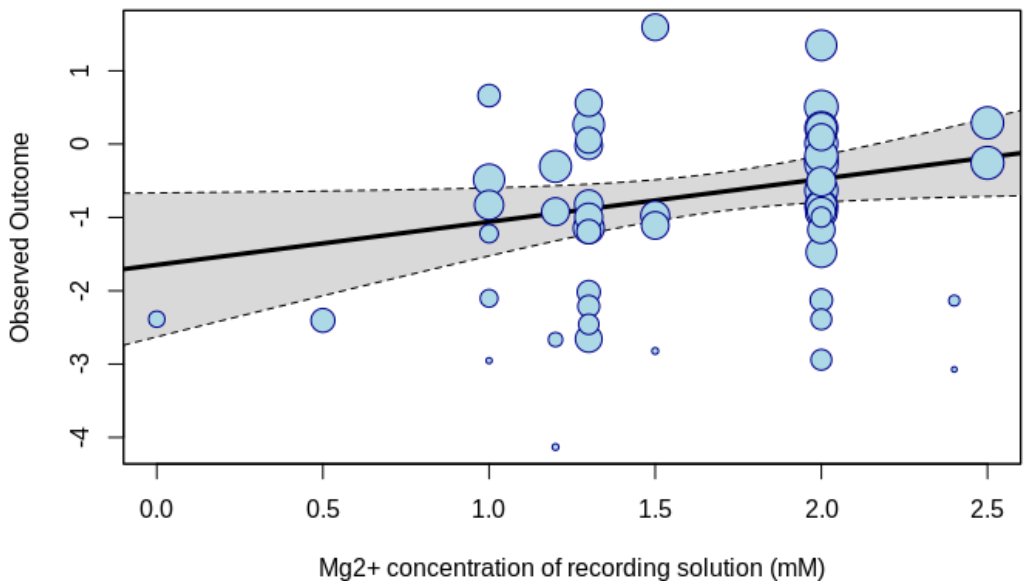


Figure 5.17: Bubble plot of meta-regression analysis between SMD effect size and magnesium concentration of recording solution in LTP experiments at 30 minutes

Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=54 comparisons.

Variable	Estimate (SE)	95% CI	P value	Tau^2	I^2	R^2
Intercept	0.62	-0.37 to 1.61	0.0097	0.60	84.25	13.8
Baseline % I/O	-0.03	-0.06 to -0.01				

Table 5.17: Univariable meta-regression of LTP outcomes at 30 minutes with baseline stimulation as percentage of I/O. N=64 comparisons

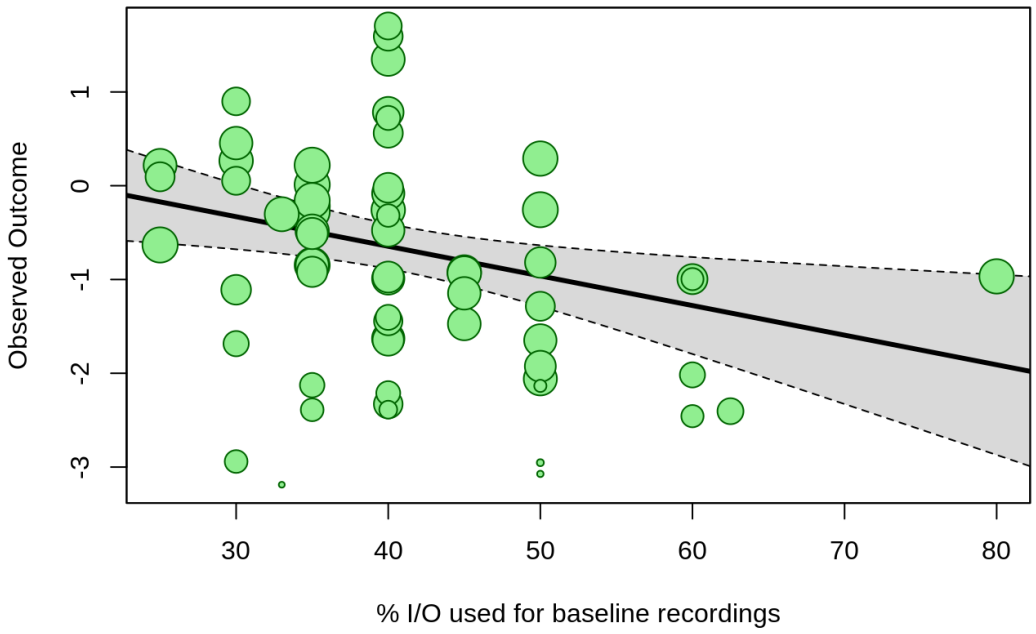


Figure 5.18: Bubble plot of meta-regression analysis between SMD effect size and % maximal I/O stimulation used for baseline recordings in LTP experiments at 30 minutes
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=64 comparisons.

Variable	Estimate (SE)	95% CI	P value	Tau^2	I^2	R^2
Intercept	-0.39	-0.76 to -0.02	0.0327	0.76	85.59	9.26
Number of stimulations	-0.002	-0.004 to 0.00				

Table 5.18: Univariable meta-regression of LTP outcomes at 30 minutes with number of stimulations
N=87 comparisons, CI =confidence intervals.

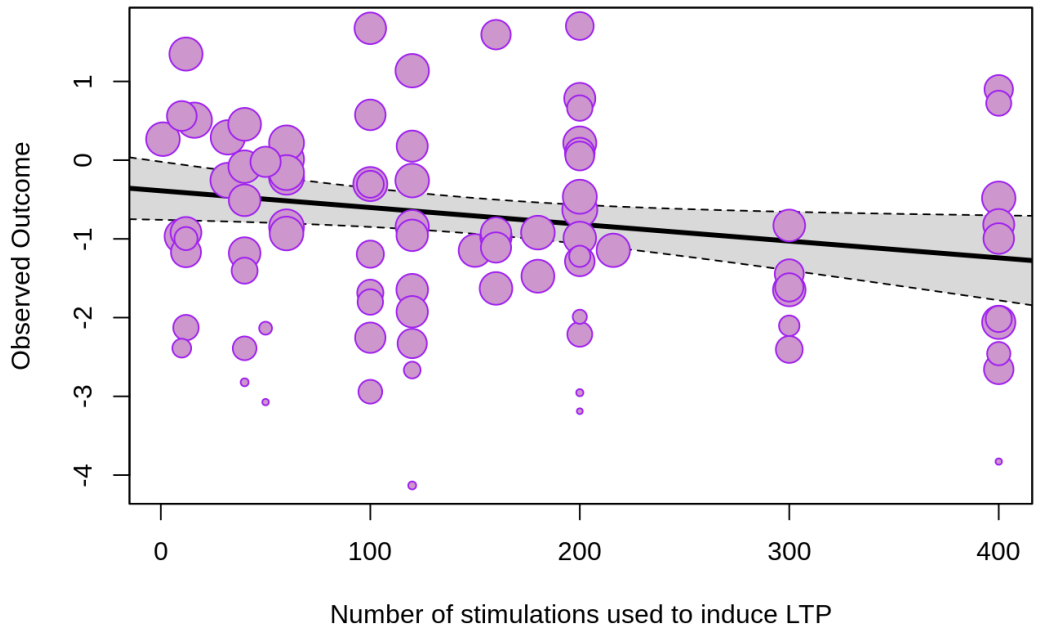


Figure 5.19: Bubble plot of meta-regression analysis between SMD effect size and number of stimulations to induce LTP at 30 minutes.
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=87 comparisons.

To build a multivariable meta-regression model, I began with transgenic model as it explained the greatest proportion of the heterogeneity. Strain and reporting of conflict of interest statements were added as significant predictors as they led to the most significant reductions in AIC. No further predictors were added to the model. Age of animals at outcome assessment also led to a significant reduction in the AIC when added to transgenic model in the meta-regression, but strain led to a better model fit so was retained in the model over age of outcome assessment. The final model explained 27.28% of the heterogeneity in the dataset.

Variable	Regression weight (β coefficient)	95% CI	P value	Tau^2	I^2	R^2
Intercept	-1.34 (0.43)	-2.2 to -0.48	0.0005	0.58	81.35	27.28
5xFAD	0.09 (0.54)	-0.98 to 1.17				
APPswe/PSEN1dE9	0.43 (0.39)	-0.35 to 1.21				
Other	0.78 (0.4)	-0.01 to 1.57				
Tg2576	1.09 (0.46)	0.18 to 2				
Strain: C57BL6/SJL	0.6 (0.48)	-0.35 to 1.54				
Strain: NR	0.09 (0.32)	-0.54 to 0.72				
Strain: Other	0.93 (0.32)	0.3 to 1.56				
COI statement: Reported	-0.49 (0.22)	-0.94 to -0.04				

Table 5.19: Multivariable meta-regression of LTP outcomes at 30 minutes
N=93 comparisons, CI =confidence intervals, COI= Conflict of interest statement.

To visualise the relationship between age and model, I also plotted bubble plots of age of outcome assessment by transgenic model (Figure 5.20). However, age did not explain a significant proportion of the heterogeneity for any of the model subgroups.

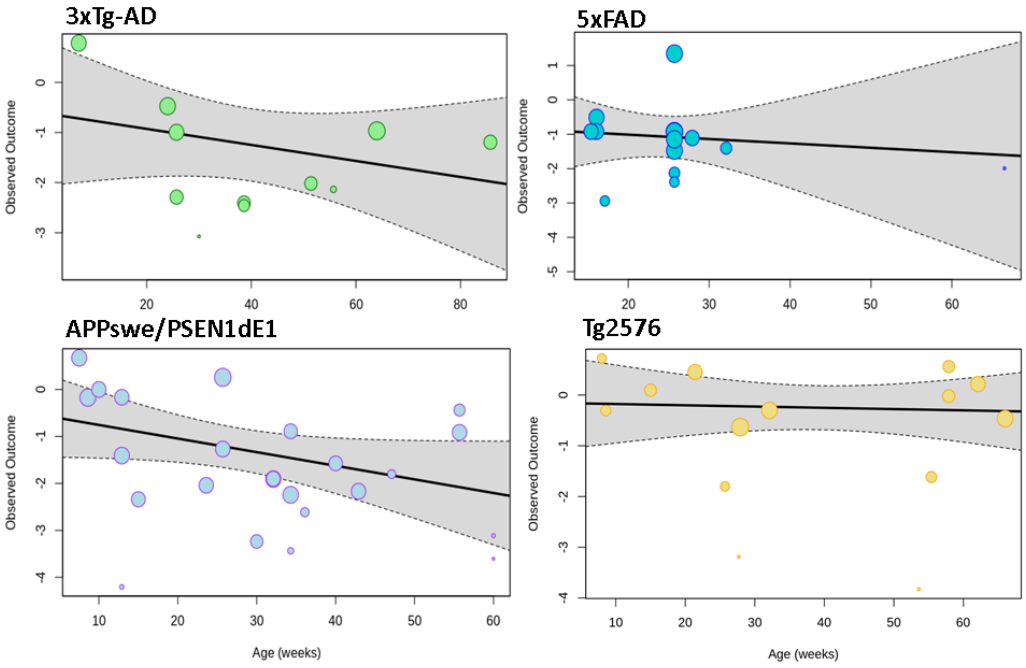


Figure 5.20: Bubble plot of meta-regression analysis between SMD effect size and age outcome measured grouped by transgenic model, at 30 minutes post-LTP.
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison.

In a post-hoc exploratory analysis, I ran repeated chi-square tests to understand the relationship between conflict of interest statements and other risk of bias reporting variables (Blinding, Animal welfare committee approval, Exclusion criteria) and study-level electrophysiology protocols (anaesthesia prior to sacrifice, recording chamber type). With Bonferroni adjustment, the p-value signifying a significant result was 0.01. I identified a significant relationship between reporting of animal welfare committee approval and conflict of interest statements (X-squared = 8.8106, df = 1, p-value = 0.003).

I converted conflict of interest reporting to a binary numeric variable (1,0) and conducted a Pearson's product-moment point-biserial correlation analysis between conflict of interest reporting and Year of publication. There was a highly significant correlation (r=0.61 (95% CI: 0.46 – 0.72), df=90, p<0.0001). Finally, I ran an additional univariable meta-regression to observe effect sizes by year and visualised the relationship in a bubble plot (Figure 5.21). Publication year explained a significant proportion of the heterogeneity, with larger reductions in LTP (reduced SMD effect sizes) over time (R^2 = 14.74%, p-value = 0.0007).

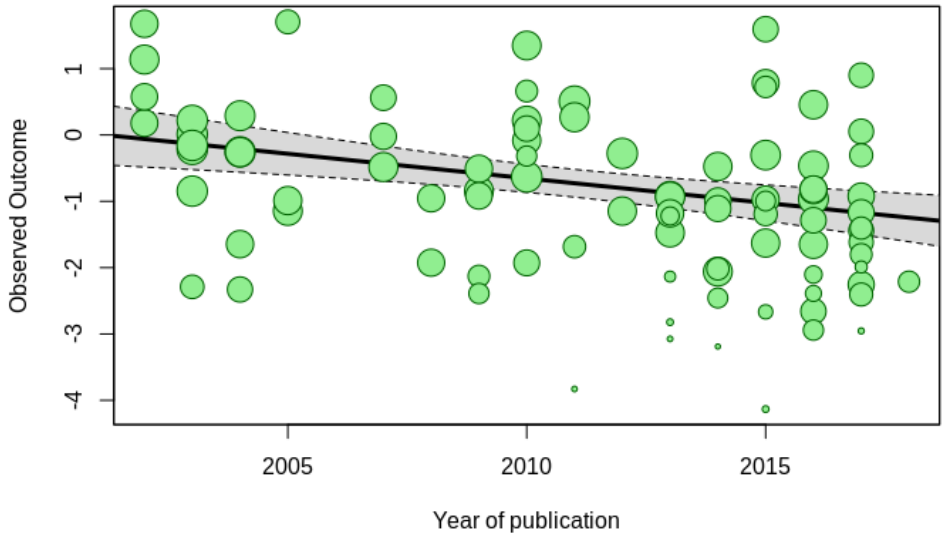


Figure 5.21: Bubble plot of meta-regression analysis between SMD effect size and year of publication at 30 minutes post-LTP.
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=92 comparisons.

2. ***Long-term potentiation in CA1 at 60 minutes***

In modelling experiments measuring LTP at a post-stimulation duration of 60 minutes, the pooled effect of transgenic modelling interventions was -0.74 (95% CI –0.98 to -0.51, n=72 comparisons), meaning that transgenic animals had reduced LTP versus wild-type controls. Heterogeneity was high (I^2 = 78.7%, Q=333.59, Tau^2=0.67). A forest plot of comparisons is shown in Figure 5.22.

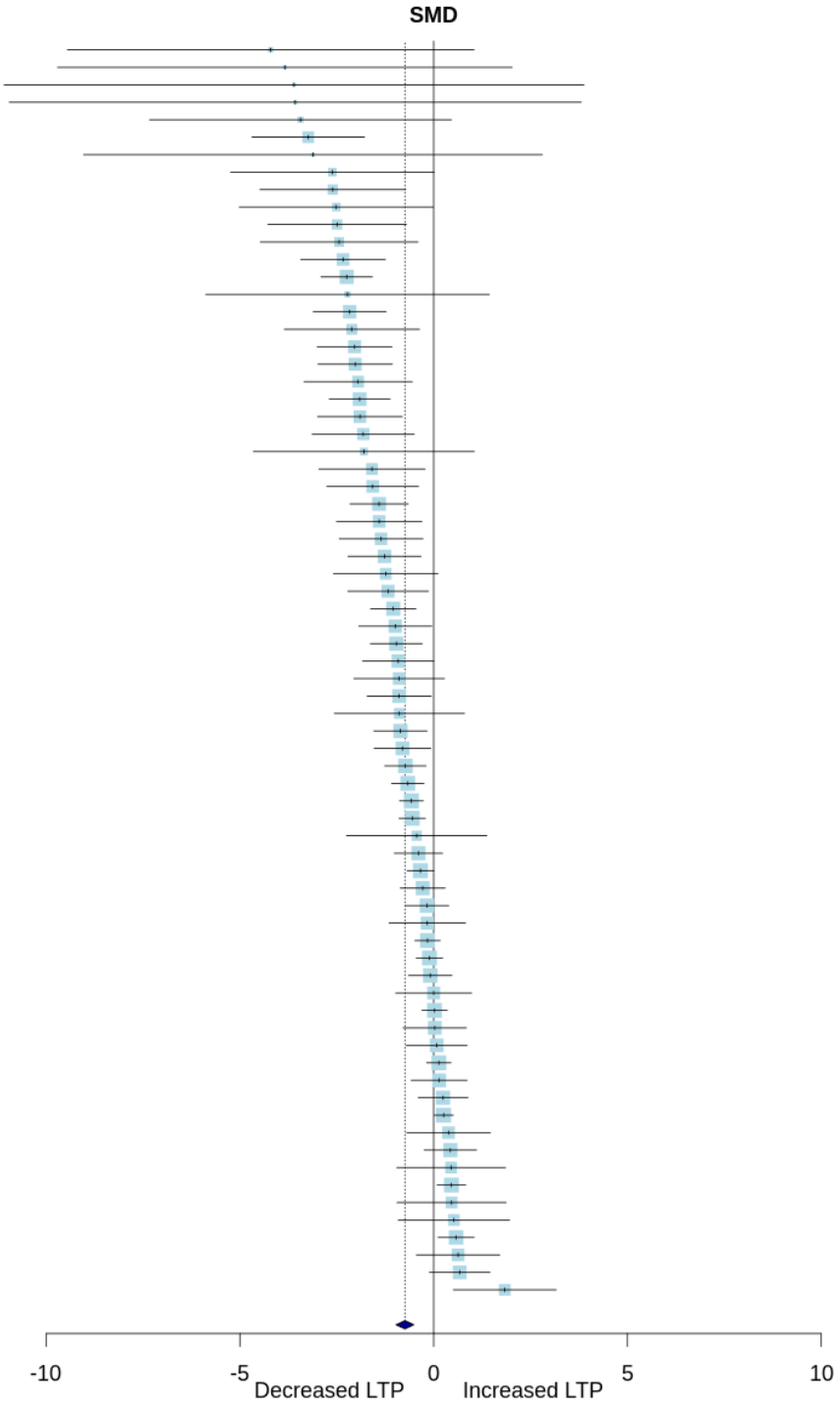


Figure 5.22: Forest plot of random effects meta-analysis of LTP after 60 minutes in modelling experiments
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. https://osf.io/5p9a2/

Several variables explained a significant proportion of the heterogeneity in the univariable analysis. (Table 5.20 – 2.24). As seen at the 30 minute post LTP induction timepoint, transgenic model and transgenic model category each explained a significant proportion of the heterogeneity (R^2 = 17.31). Again, these results suggest LTP was less impaired in the Tg2576 model and APP models in general. Reporting of a conflict of interest statement and reporting of an approval statement from an animal welfare committee were associated reduced LTP (lower SMD estimates). Reporting of the use of anaesthesia prior to animal sacrifice for electrophysiology experiments was associated with an increase in LTP. A relationship was observed (as at the 30 minute timepoint) with magnesium concentration of recording solution, where increased magnesium concentration was associated with increased estimates of LTP (increased SMD estimates). Every 1mM increase in magnesium concentration was associated with an SMD increase of 0.67 (Table 5.24, Figure 5.23). For the main analysis, I retained the unusual observation where an experiment had not used any magnesium (0mM magnesium) for the recording solution. However, to assess the potential impact of this, I performed a post-hoc leave-one-out univariable analysis omitting that datapoint. In this analysis, a 1mM increase in magrnisum was associated with an SMD increase of 0.62 and magenisum concentration explained a slightly smaller proportion of the heterogeneity (R^2=20.92).

Of all variables examined in the univariable meta-regression analysis, magnesium concentration explained the greatest proportion of the heterogeneity (R^2 = 23.00) but could not be taken forward into the multivariable meta-regression model as only 41/72 comparisons reported this information. I did not conduct multivariable meta-regression as no additional variables (when added to transgenic model category) significantly improved the model fit.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APPSwe/PSEN1de9	-1.26	-1.67 to -0.85	0.0054	0.56	81.78	17.31
ModelOther	-1.94	-1.08 to -0.07				
ModelTg2576	-0.18	-0.89 to 0.54				

Table 5.20: Univariable meta-regression of LTP outcomes at 60 minutes with model
N=72 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APP	-0.26	-0.62 to 0.09	0.0038	0.56	82.04	17.54
APP/PS1	-1.08	-1.56 to -0.61				
APP/PS1/MAPT	-0.78	-1.57 to 0.02				

Table 5.21: Univariable meta-regression of LTP outcomes at 60 minutes with model category
N=72 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
COI statement: Not reported	-0.38	-0.71 to -0.06	0.0039	0.58	82.73	14.59
COI statement: Reported	-1.06	-1.52 to -0.61				

Table 5.22: Univariable meta-regression of LTP outcomes at 60 minutes with reporting of conflict of interest
N=72 comparisons, CI =confidence intervals, COI= Conflict of interest

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Anaesthesia prior to sacrifice: Not reported	-0.97	-1.28 to -0.65	0.0035	0.62	83.42	8.42
Anaesthesia prior to sacrifice: Yes	-0.46	-0.93 to -0.01				

Table 5.23: Univariable meta-regression of LTP outcomes at 60 minutes with anaesthesia prior to sacrifice
N=72 comparisons, CI =confidence intervals

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Intercept	-1.69	-2.63 to -0.75	0.0124	0.24	72.16	23.00
Magnesium concentration of recording solution	0.67	0.15 to 1.19				

Table 5.24: Univariable meta-regression of LTP outcomes at 60 minutes with magnesium concentration of recording solution
N=41 comparisons, CI =confidence intervals.

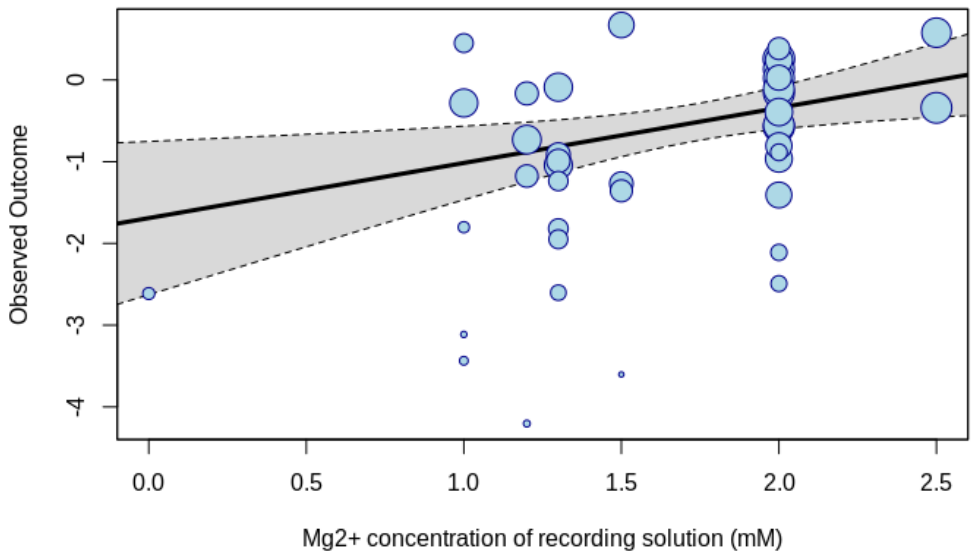


Figure 5.23: Bubble plot of meta-regression analysis between SMD effect size and magnesium concentration of recording solution in LTP experiments at 60 minutes.
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=41 comparisons.

3. *Maximum input/output relationship at CA1*

In modelling experiments measuring the input/output relationship (basal synaptic transmission) in hippocampal slices in the CA1, the pooled effect of transgenic modelling interventions was -0.60 SMD (95% CI –0.85 to -0.35, n=63 comparisons). Overall, transgenic animals had reduced basal synaptic transmission versus wild-type controls. Heterogeneity was high (I^2 = 82.1%, Q=347.15, Tau^2=0.73). A forest plot of comparisons is shown in Figure 5.24.

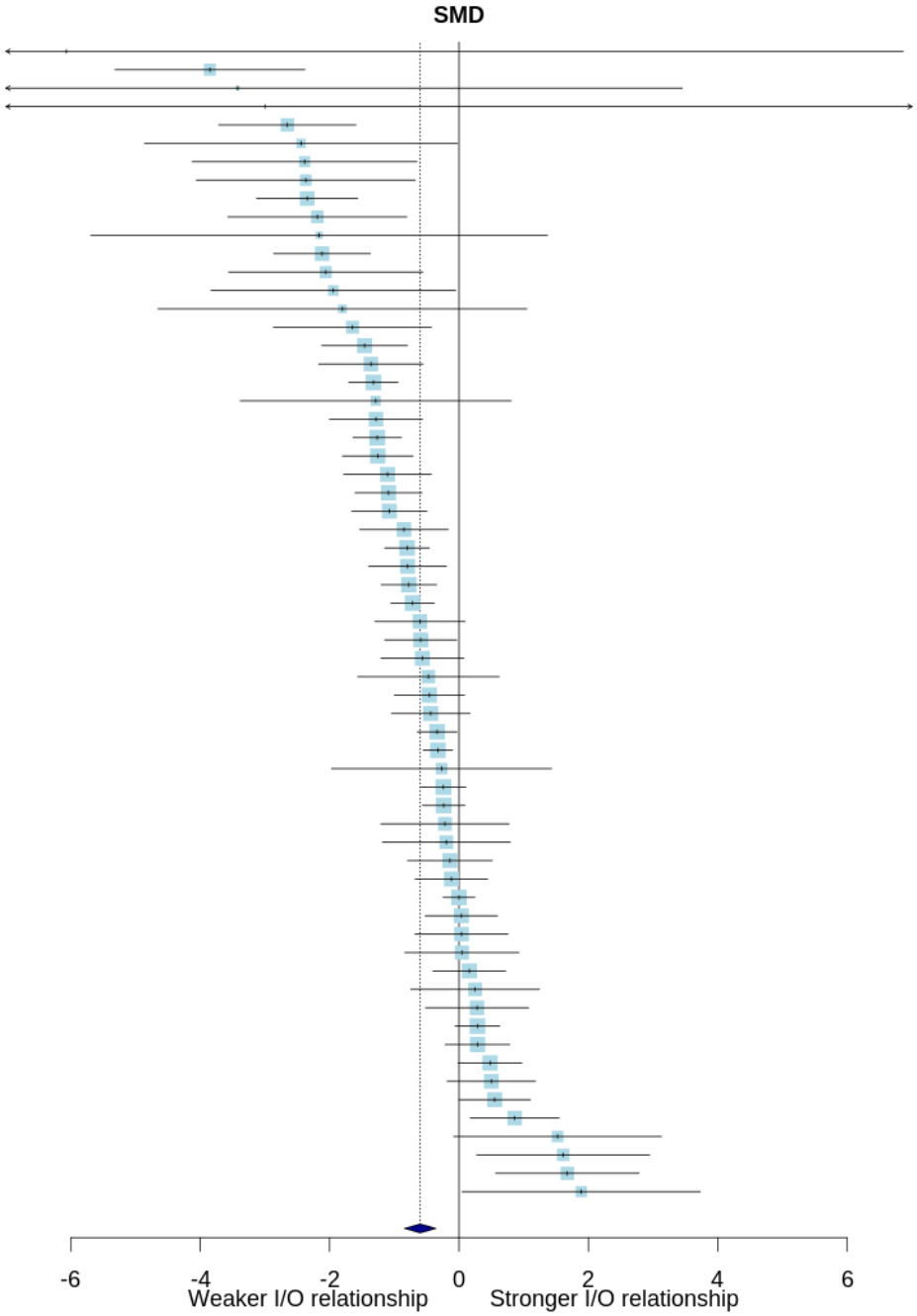


Figure 5.24: Forest plot of random effects meta-analysis of input/output relationship in modelling experiments
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate Full size figure available at: https://osf.io/vrfum/

In the univariable analysis, only transgenic model explained a significant proportion of the heterogeneity (Table 5.25). However, it is difficult to interpret this as most models were grouped into an "Other" category due to low sample sizes. In the model building exercise, reporting of a conflict of interest statement was associated with a significant reduction in SMD estimates (reduced LTP).

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APPswe/PSEN1dE9	0.05	-0.41 to 0.52	0.00019	0.61	87.32	16.66
Model: Other	-0.83	-1.38 to -0.29				

Table 5.25: Univariable meta-regression of input/output outcomes with transgenic model N=63 comparisons, CI =confidence intervals.

Variable	Regression weight (β coefficient)	95% CI	P value	Tau2	I2	R^2
(Intercept)	0.45 (0.27)	-0.09 to 0.99	0.1041	0.55	86.28	23.97
Model: Other	-1.14 (0.28)	-1.7 to -0.58	1e-04			
Conflict of interest statement: Reported	-0.69 (0.27)	-1.24 to -0.15	0.0134			

Table 5.26: Multi-variable meta-regression of input/output outcomes N=63 comparisons, overall p-value= 0.0004, CI =confidence intervals.

5.4.17 Relationship to cognition (MWM)

Few studies reported MWM outcomes in the same cohort of animals as electrophysiology measurements. In publications included in the largest dataset (LTP at 30 minutes post-induction), I identified 34 comparisons across 12 studies which either reported MWM outcomes in the same cohort of animals or, where unclear if the same cohort was used, measured MWM outcomes prior to electrophysiology outcomes (i.e. rodents were the same age or younger when MWM outcomes were measured). Effect sizes from acquisition (n=19, R^2=0%,

p=0.72) or probe (n=15, R^2=26.75, p=0.08) phase MWM outcomes did not explain a significant proportion of the heterogeneity seen in LTP. The relationship between MWM and LTP outcomes are shown in Figure 5.25.

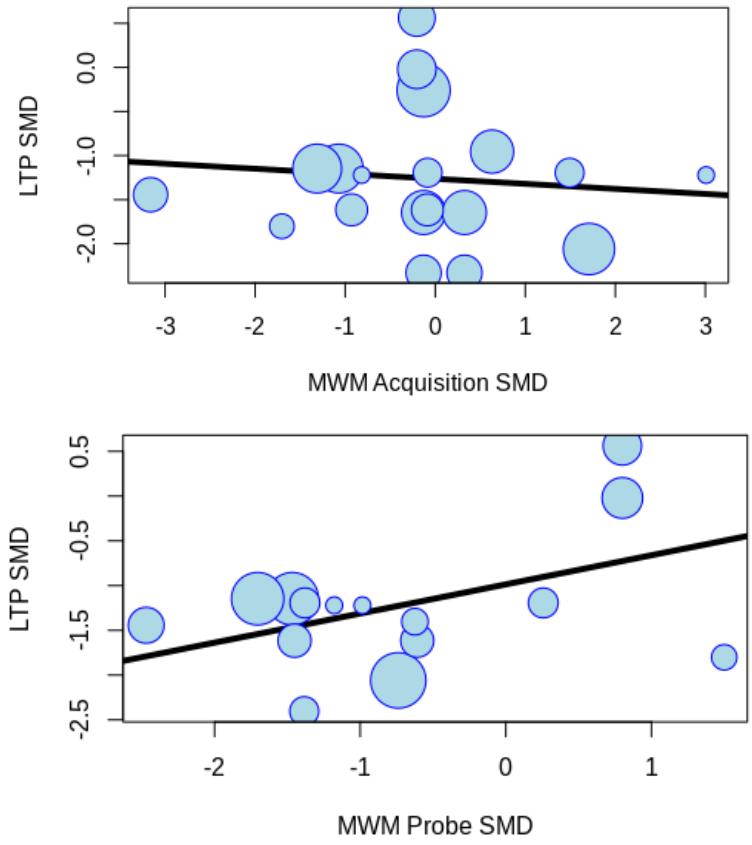


Figure 5.25: Bubble plots showing meta-regression analysis between MWM outcomes and LTP (at 30 minutes duration)

Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. N=19 acquisition comparisons, N=15 probe comparisons.

5.4.18 Publication bias

1. ***Long term potentiation in CA1 at 30 minutes***

Visual inspection of the funnel plot (Figure 5.26) indicated asymmetry. Egger's regression also suggested funnel plot asymmetry for LTP outcomes (t = -2.101, df = 90, p-value = 0.038). These results are indicative of small study effects, which may suggest the presence of publication bias. Using trim and fill analysis, the inclusion of 27 theoretical missing studies decreased the estimate of synaptic plasticity deficits in transgenic AD models by 52.8% to -0.34 (95% CI -0.58 to -0.10)

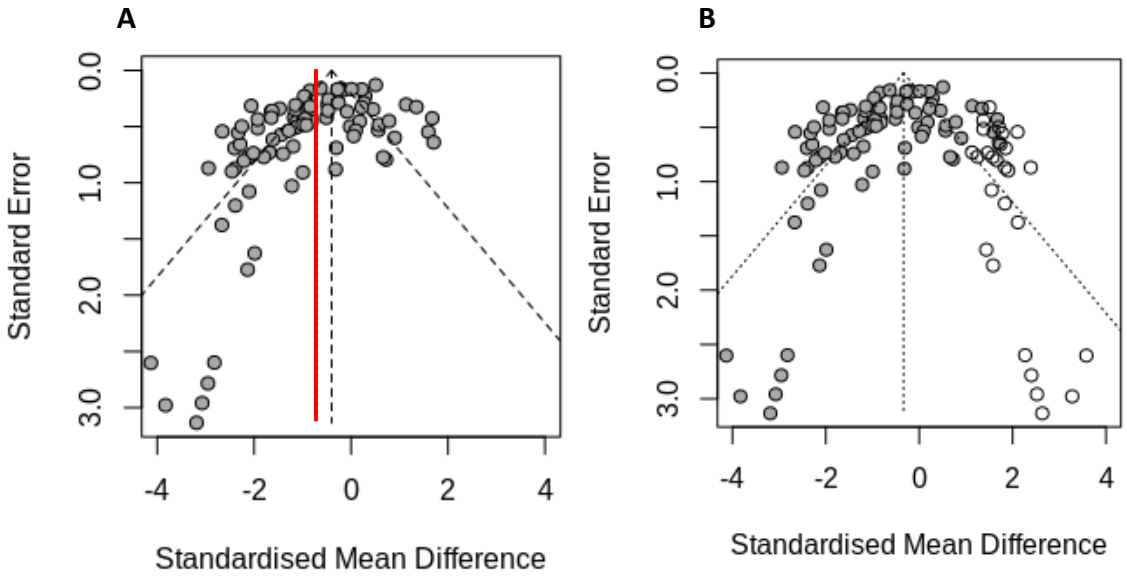
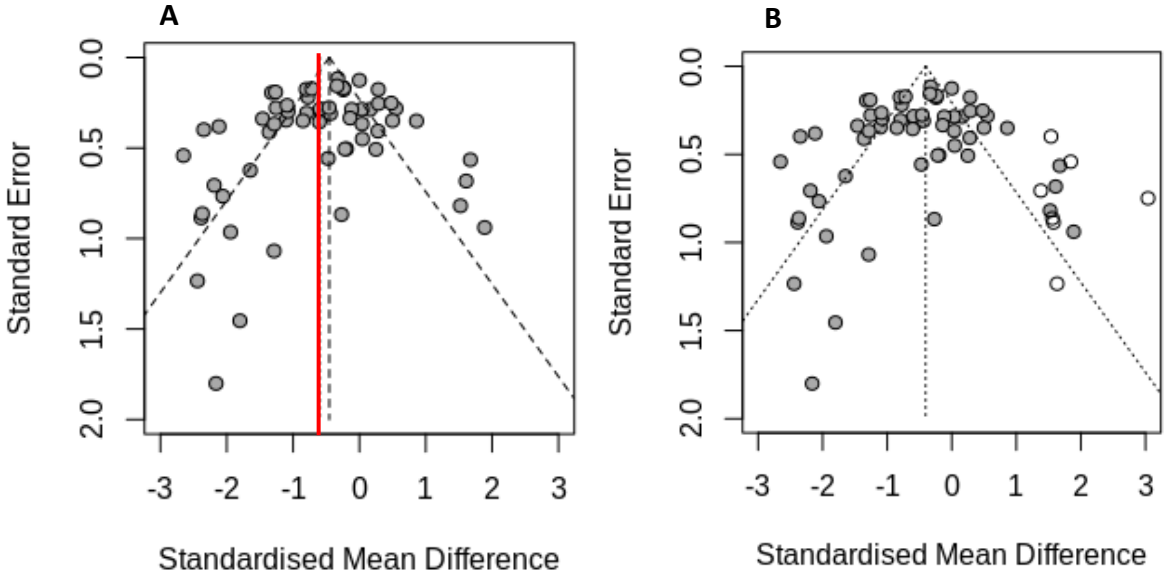


Figure 5.26: Funnel plots of effect sizes for LTP at 30 minutes
A: The red solid line indicates global estimate of effect. (B) Theoretical missing experiments are shown as unfilled circles. Filled circles represent reported experiments. The dotted line represents adjusted global effect size including missing experiments.

2. ***Maximum input/output relationship at CA1***

Eggers's regression did not indicate funnel plot asymmetry (t = 0.21, df = 61, p-value = 0.83). On visual inspection of funnel plots, there appears to be slight asymmetry towards the left hand side. Trim and fill analysis added 10 theoretically missing studies and decreased the effect size estimate to -0.40 SMD (95% CI: -0.68 to -0.13). Overall, these results indicate that there may be a small amount of publication bias, but it does not have a large effect in this dataset.



5.5 Discussion

5.5.1 Overview of findings

This SR and meta-analysis of *in vitro* electrophysiology outcomes measured in APP transgenic models summarised the literature across 166 publications describing modelling interventions and treatment interventions. I did not conduct a meta-analysis of treatment effects as only one treatment was evaluated in more than one publication. This lack of independent verification may be indicative of the wider

reproducibility crisis across the field. More incentives, such as funding specifically for replication efforts, are required to enable researchers to verify the reported impact of therapeutic interventions.

Overall, the meta-analysis of transgenic models suggested that transgenic AD mice have reduced measurements of LTP (synaptic plasticity) and I/O relationships (baseline synaptic strength) versus wild-type controls. The heterogeneity observed in across experiments was high.

5.5.2 Modelling of Alzheimer's disease in APP models

Across outcomes, transgenic animal model explained a significant proportion of the heterogeneity. In LTP experiments at both 30 and 60 minutes post-stimulation, APP models and specifically experiments in Tg2576 mice had higher SMD estimates, suggesting that synaptic impairments in this model were less evident. However, looking at the distribution of age outcomes were measured (Figure 5.5), this could be influenced by many Tg2576 experiments measuring animals early in the development of AD pathology, and prior to amyloid deposition. In LTP experiments measured at 30 minutes, experiments in the 3xTg-AD model had the greatest reduction in LTP. At the 60 minute LTP timepoint, experiments APP/PS1 models had the largest reduction in SMD effect size. Together, these observations suggest that there is less evidence of LTP impairments in single APP models, while in APP/PS1 and APP/PS1/MAPT models, impairments are more obvious. However, it is important to note that within each model, there was variation (see Figure 5.16). Some experiments within the same transgenic model showed a worsening of LTP, while others seemed to show enhancement of LTP. Interestingly, there appears to be no ceiling effect of AD modelling on synaptic plasticity impairment. Each LTP forest plot (Figure 5.15-16, 5.22) shows that the largest reductions in LTP we observed are associated with the largest variability, due to the small sample sizes used within those experiments.

The sex of animals used did not explain a significant proportion of the heterogeneity observed between studies for any outcomes, however, the lack of reporting and tendency to use only male animals makes this hard to adequately assess.

The age of animals at time of outcome assessment did not explain a significant proportion of the heterogeneity for any outcome. Previous work suggests there is a age-dependent relationship with synaptic plasticity in both wild type (Lynch, 2004; Rogers et al., 2017) and AD mice (Gengler, Hamilton, & Hölscher, 2010) but this was not observed in the meta-regression analysis. However, recent findings have also shown a recovery of LTP in older mice (Huh et al., 2016). It may be that the high heterogeneity we observed may be masking subtle biological relationships. There is a clear need to reach a better understanding of how age impacts upon both wild-type LTP measures and measurements in transgenic AD models.

5.5.3 Pseudoreplication and sample size

A key finding of this review is that a substantial proportion of experiments in this literature (53% of modelling experiments included in this review) did not report the sample size i.e. number of animals for at least one experiment. Pseudoreplication occurs when the sample size used for analysis is artificially inflated, which can lead to exaggerated estimates of effect (Lazic, 2010; Lazic, Clarke-Williams, & Munafò, 2018) unless appropriate statistical methods e.g. generalized linear mixed-effects models (Millar & Anderson, 2004) are used. The experimental unit for the modelling experiments I included in this review is the experimental animal. An intervention, in this case a transgenic modelling intervention, has been applied to individual animals and it is the difference between animals with and without this intervention that experiments aim to measure. This logic stands even for *in vitro/ex vivo* measurements where organ or tissue is taken from an animal and outcomes measured on that sample. Despite this, a huge number of experiments report only the number of slices or are unclear about whether the N reported was slices or

experimental animals. To avoid pseudoreplication impacting on the results of the meta-analysis, I chose to only include studies which had specifically mentioned the number of animals. This widespread issue limited our ability to utilise evidence in the meta-analysis and limited our power to identify important sources of heterogeneity. In addition, only 4/166 publications reported a power calculation to determine an adequate sample size. Together, these omissions raise serious questions about the strength of existing evidence in this field and whether the extent of impairments in synaptic functioning or the efficacy of treatments to reverse these deficits may have been overstated.

5.5.4 Internal validity of experiments

Reporting of study quality and measures to reduce the risk of bias were moderate to poor. Only four studies (2.4%) reported a sample size calculation to ensure their experiments were adequately powered. Further, just under a third of studies (31.1%) reported that outcome assessors were blinded to experimental groups. Most studies reported a conflict of interest statement (62.3%) and most had attained welfare committee approval (69.9%). It is not possible to assign animals randomly to the transgenic model (this is done naturally). Therefore, randomisation and allocation concealment were only applicable for publications reporting a treatment comparison. Of these, only 18.9% reported randomising animals to groups and 10.8% reported that group allocation was concealed.

At a duration of 30 minutes post stimulation, in the univariable meta-regression analysis, publications reporting blinded outcome assessment had less evidence of LTP impairment compared to non-blinded studies. Interestingly, for both LTP timepoints, reporting of conflict of interest was associated with a greater reduction of LTP. Reporting of a conflict of interest statement also explained a significant proportion of the heterogeneity in the multivariable meta-regression model for LTP outcomes at 30 minutes. In a post-hoc analysis, I identified that conflict of interest

reporting was highly correlated to year of publication, and that year explained a significant proportion of the heterogeneity (R^2 = 14.74%). Increasing evidence of LTP impairment over time could be down to several factors and it is not possible to disentangle these with any certainty in this analysis. It could be hypothesised that experimental protocols have improved incrementally over time, and now the disparity between wild type and control animals is clearer. It may also be that publication bias or selective outcome reporting has increased over time and there is a greater tendency to publish studies which show larger decreases in LTP and other indicators of synaptic function in transgenic AD models.

5.5.5 Publication bias

For LTP experiments, I found evidence of substantial publication bias which suggested global effect sizes may have been overstated by 53%. There was mixed evidence of publication bias for I/O outcomes. Publication bias is a widespread issue across preclinical research (Korevaar, Hooft, & Ter Riet, 2011). If research showing that LTP was not reduced in a transgenic AD model is more likely to remain unpublished, this complicates our ability to draw conclusions about the true model phenotype and severity of impairment. Greater uptake of initiatives such as Registered Reports (in which the background and methodology of a study are submitted to a nominated journal and peer reviewed prior to data collection and analysis) may encourage the publication of more research data, irrespective of results.

5.5.6 Impact of electrophysiology protocols on modelling experiments

The impact of varying electrophysiological protocols was often hard to determine due to incomplete reporting. An interesting finding which emerged from this analysis was that magnesium concentration of the recording solution explained a significant proportion of the heterogeneity in both LTP datasets. Increasing

magnesium concentrations were associated with improvements in synaptic plasticity measures (increased LTP in transgenic models). Higher magnesium concentrations could potentially mask the impact of impairments in AD models, while low concentrations could emphasise the LTP differences. These results suggest a disparity in magnesium responsiveness between transgenic and wildtype animals, which may warrant further investigation. Given that magnesium ions are critical to the development of LTP, it would be useful to understand the underlying mechanism for this observation in animal models of AD pathology.

Unfortunately, I could not explore this effect further in the multi-variable meta-regression as there was too much data missing to perform the analysis. Furthermore, anaesthesia use before sacrifice was also associated with an improvement in outcome at 60 minutes post-LTP, where modelling experiments reporting this measure had higher estimates of effect (increased LTP) and less evidence of impairment. However, it was unclear for most papers whether anaesthesia had indeed not been used or whether details of anaesthesia were simply omitted.

5.5.7 Limitations

This work has several limitations. All data summarised here were collected by a single reviewer, either myself or my summer student. I sense-checked the dataset throughout data cleaning and analysis, and corrected errors where feasible. Despite these efforts, it is likely that some errors remain and could have an impact upon the results. In addition, there were some deviations from the study protocol. Firstly, due to variation in the way LTP is analysed throughout the field, we had intended to extract LTP at each 5 minute timepoint from figures and calculate area under the curve up to every 30 minute timepoint e.g. 30 minutes, 60 minutes, 90 minutes, 120 minutes. In practice, this was an overly time-consuming approach and there were many graphs where the data was unreadable for certain timepoints, or where the

time bins were 10 or 15 minutes rather than 5. During data extraction, we therefore decided to extract only the 30, 60, and 90 minute timepoints (where possible). Very few publications had data beyond 60 minutes post LTP stimulation, which meant that we could not compare early and late-LTP phases as originally hoped. Furthermore, due to poor reporting of true sample size, a large proportion of studies were omitted from the modelling meta-analysis, limiting our ability to understand the impact each variable has on heterogeneity across different outcomes. We were unable to conduct analysis of LTD outcomes for example, as there were so few remaining comparisons after removing those with no sample size. Furthermore, due to a lower number of usable comparisons than expected, I was unable to conduct draw meaningful conclusions from the regression between Morris water maze and electrophysiological outcomes. This was compounded by a lack of clarity in whether the same animals had been used for both behavioural and electrophysiological outcomes in most experiments. Furthermore, I was not able to consider the use of different blockers and inhibitors of synaptic transmission in any meta-analysis due to the low number of comparisons utilising each different inhibitor. The use of blockers is likely to explain at least some of the variation within a small number of comparisons. Finally, I first prioritised studies with more than 3 matches with the electrophysiology regular expression, but during data extraction I added in all studies with over 1 mention of electrophysiology. This approach allowed me to ensure I did not miss any relevant studies and contributed to my work to validate regular expression approaches to streamline the SR process (Chapter 3).

5.5.8 Future directions

To gain a better understanding of the heterogeneity in experiments reporting *in vitro* electrophysiological outcomes in hippocampal slices, greater transparency in the reporting of experimental design is required. If the conduct of experiments is

transparent and key parameters are reported, we will be able to gain a clearer profile of the evidence for synaptic dysfunctions in each transgenic model. Furthermore, research improvement efforts are needed within the slice electrophysiology community to ensure that experiments use an adequate sample size and that statistical tests based on correct experimental unit for the aims of the experiment. This review has identified many gaps in the existing literature. Relatively few publications measure LTD in transgenic AD models and most experiments were conducted in young or adult male mice, with fewer experiments in aged mice or female mice. Where LTP and/or LTD were measured, this was often for short periods of time following induction. It would be useful to gain a clearer understanding of the change in LTP over time. Most research efforts have also been focussed on the CA1-CA3 pathway, whereas other hippocampal pathways have been less frequently investigated. Lastly, although this work was focused on *in vitro* electrophysiological measurements of *in vivo* interventions (both transgenic modelling and drug interventions), it would be beneficial in future work to determine the relationship between *in vivo* and *in vitro* LTP measurements, and identify differences observed between testing a drug intervention *in vivo* and *in vitro.*

Once the annotated data collected during this review has been fully reconciled, I will make the dataset publicly available and interrogatable by researchers to guide future research questions.

CHAPTER 6: A CROWDSOURCED SYSTEMATIC REVIEW AND META-ANALYSIS OF THE OPEN FIELD TEST IN TRANSGENIC ALZHEIMER'S MOUSE MODELS

6.1 Chapter Introduction

In this chapter, I describe the methodology for, and results of, a SR and meta-analysis project which was conducted in collaboration with the European Quality in Preclinical Data (EQIPD) consortium. I designed the study protocol, developed training, recruited reviewers, and managed the project. A detailed protocol was pre-registered prior to the data extraction phase and deposited on the open science framework (Appendix C6.1). This was an ambitious SR project with over 700 potentially relevant publications. Dual-data extraction has almost been completed, but we do not currently have the resources required to co-reconcile the resulting data by a third independent reviewer. I have instead chosen to summarise and analyse a sub-set of the modelling dataset focused on open-field test behaviour in the most used amyloid precursor protein (APP) mouse models of Alzheimer's disease (See Chapter 1 for details on transgenic APP models). A team of trained reviewers contributed to the data extraction (see Acknowledgements). To reflect the collaborative nature of this work, I therefore use "we" when referring collectively to the team of reviewers who worked on the project.

6.2 Background

6.2.1 The history of the open field test

The open field test (OFT) was first described over 80 years ago (Hall, 1934) as a paradigm to assess emotional elimination (defecation) in rats. Since then, the OFT has been used extensively to assess multiple aspects of spontaneous rodent behaviour with relative ease and efficiency including anxiety-like behaviour,

locomotor activity, and exploration, (Seibenhener & Wooten, 2015; Tanaka, Young, Halberstadt, Masten, & Geyer, 2012) and remains one of the most commonly reported behavioural tests for both rats and mice (Stanford, 2007).

During the test, rodents are placed in a novel open space, typically a circular or square arena, surrounded by a wall to prevent escape. The paradigm may be an attractive option for behavioural characterisation due to its relative simplicity, non-invasiveness, and range of potential measurements for phenotypic characterisation.

6.2.2 Measuring locomotor activity and exploration

The most reported OFT measure is total distance travelled i.e., the path travelled (in centimetres or metres) by rodents across the arena. Horizontal activity is a similar measure, but is typically recorded via beam breaks which may also represent body movements unrelated to ambulatory activity such as rearing or head movements. Vertical activity, mainly indexed by rearing behaviour where the animal stands vertically on its hind legs, is another measure of general locomotor activity (Wexler, Benjamini, & Golani, 2018).

6.2.3 Measuring anxiety and emotionality

The traditional definition of an emotional animal in the open field test is quantified by low activity and high defecation (Denenberg, 1969). When exposed to the open field (a stressor), the rationale is that the autonomic nervous system will be triggered and defecation will occur, as noted in the paradigm's conception (Hall, 1934). Additionally, immobility or "freezing" in response to noxious stimuli may have an adaptive benefit in that it reduces the likelihood of being noticed by predators. Rodents are naturally averse to open, exposed spaces. Thigmotaxis behaviours i.e., the tendency to remain in sheltered areas can be measured in the open field test by the time spent in or entries to corners and peripheral sections of

the arena. The time an animal spends in the centre area (where a reduction indicates greater anxiety) is sensitive to most traditional anxiolytics, but remained unchanged with the newer generation anxiolytic medications such as selective serotonin re-uptake inhibitors (Prut & Belzung, 2003). Self-grooming or washing behaviours are seen as an adaptive response to reduce anxiety in stressful situations (Kametani, 1988). Increased grooming behaviours during the open field test have been identified in animals under increased stress (Moyaho & Valencia, 2002). Further, as slow self-grooming can also occur in non-stressful situations, self-grooming accompanied by frequent "interrupted" grooming bouts (Kalueff & Tuohimaa, 2004) may be a more sensitive measure to detect anxiety.

Rearing behaviour also has a stress-sensitive dimension, with stressed animals rearing less frequently (Sturman, Germain, & Bohacek, 2018). In addition to vertical activity, the frequency and duration of rearing behaviours may therefore also represent an additional measure of anxiety. Anxiety-like outcomes in the open field have been successfully linked to other behavioural measures of anxiety such as the elevated plus maze (Carola, D'Olimpio, Brunamonti, Mangia, & Renzi, 2002).

6.2.4 Separation of OFT measures

It remains difficult to disentangle the biological meaning OFT outcomes intended to measure anxiety or activity. If there is reduced activity in the OFT, this will impact upon anxiety-like measures, such as distance travelled in the centre area or central zone entries. Conversely, there are a number of underlying facets which may affect how much an animal moves throughout the arena. As described in a detailed critical review of the paradigm (Walsh & Cummins, 1976), the evocation of animal behaviours is dependent on several factors, including (1) removal from their home environment, (2) transference to the OFT arena, and (3) exposure to a novel test environment, and (4) prior experience of similar test situations. Such factors will likely influence anxiety levels and drive the animal to explore the arena

(Seibenhener & Wooten, 2015). As others have pointed out (Crawley et al., 1997), no specific OFT parameter captures a unique trait (e.g., anxiety), all measures are at least partially inter-dependent. Despite these limitations, to provide an overview of how OFT results are typically interpreted, I have operationalised commonly used OFT outcomes and their interpretations in Table 6.1.

Behaviour	Dimension	Interpretation
Exploratory Ambulation	Activity	Motor deficits signalled by: ↓ Distance travelled
Horizontal activity	Activity	Motor deficits signalled by: ↓ Horizontal activity
Vertical Activity	Activity Anxiety	Motor deficits signalled by: ↓ Vertical activity ↓ Rearing frequency ↓ Rearing duration Anxiety deficits signalled by: ↓ Vertical activity ↓ Rearing frequency ↓ Rearing duration
Thigmostaxis	Anxiety	Increased stress/anxiety signalled by: ↓ time in central zone ↓ distance travelled in central zone ↓ activity in central zone ↓ centre:periphery ratio (acitivity, distance, duration)
Grooming	Anxiety	Increased stress/anxiety signalled by: ↑ grooming duration ↑ grooming episode frequency ↑ grooming bout transitions ↓ latency period to groom
Defecation	Anxiety	Increased stress/anxiety signalled by: ↑ defecation

Table 6.1: Operationalised OFT measures

6.2.5 Open field test measurements in preclinical Alzheimer's models

The OFT has been widely employed across the preclinical AD literature. Performance on tasks measuring cognition may also be affected by hypo/hyperactivity (Rodgers, Born, Das, & Jankowsky, 2012), and so it remains important to characterise multiple aspects of behavioural phenotypes.

Furthermore, there is evidence that motor deficits occur early in human AD pathology (Goldman, Baty, Buckles, Sahrmann, & Morris, 1999) and that these are related to the extent and progression of cognitive decline that patients experience (Buchman & Bennett, 2011). Anxiety is also a prominent and frequently reported symptom in AD patients (Aalten et al., 2003; Serra et al., 2010). However, the overall picture is complex. Patients may be inhibited (anxious), apathetic, or present with disinhibition and agitation related to frontal lobe dysfunction (Lyketsos et al., 2002; Senanarong et al., 2004). Similarly, conflicting findings are common between, and within, different AD rodent models, but it has been argued that a change in either direction is relevant. It has been hypothesised that due to a decrease in motor activity with age, hyperactivity may be more obvious in older mice and hypoactivity (reduced activity) may be more obvious in younger mice (Lalonde, Fukuchi, & Strazielle, 2012).

6.2.6 Reproducibility and validity of the open field test

"It is obvious in reviewing this area that many of the all-too-numerous discrepancies, failures of replications, and contradictions stem from preventable methodological and experimental design causes."
- (Walsh & Cummins, 1976)

Despite the OFT's continued popularity (Stanford, 2007), concerns have been raised repeatedly about the effects of experimental design on the reproducibility of findings. Perhaps unsurprisingly, light intensity (Godsil & Fanselow, 2004) has been found to affect the extent of anxiety-like behaviour in the arena, with brightly lit areas associated with higher thigmotaxis. Some authors have also suggested there is a short adaptation period to the open field, followed by a longer period which more adequately reflects the true behaviours of the rodents tested. In one study (Fonio, Benjamini, & Golani, 2012), experimenters found that measuring anxiety over a short and long term resulted in opposing results. This calls into question

whether the typically short durations that open field behaviours are measured for is capturing trait behaviour or merely a reaction to novelty. For repeated OFT measurements, several experiments have identified an effect of trial day (Asano, 1986; Bond & Di Giusto, 1977). Furthermore, like other behavioural tests, OFT measures are sensitive to sex (Knight et al., 2021), and strain-specific differences (Mandillo et al., 2008; Mathis, Paul, & Crawley, 1994).

Crucially, many of these aspects and others - including the laboratory environment, animal husbandry procedures, the size and shape of the arena, and whether or not animals were habituated to the arena - were recognised over 40 years ago (Walsh & Cummins, 1976). However, there is no indication that the recommendations set out at that time have been adopted (Spruijt, Peters, de Heer, Pothuizen, & van der Harst, 2014).

6.3 Methods

6.3.1 Study identification

Publications with experiments using the OFT were identified within the larger preclinical AD dataset (n = 22,375 publications) described in detail in Chapter 2. I developed a regular expression dictionary (Appendix C6.1) to identify OFT mentions within the full-text of the publications and used the AutoAnnotation R package (Liao, 2017) to count the frequency of OFT terms within each publication. I prioritised studies in which the frequency of terms from the OFT dictionary was greater than three, as these were deemed highly likely to measure OFT outcomes, as opposed to just mentioning the paradigm in the introduction or reference section.

6.3.2 Pilot study

Using the first iteration of the OFT regex identified 927 publications with a frequency of greater than three for the OFT dictionary. From this sample, 10 papers were selected at random to be included in the pilot study. An XML Endnote library file with citations of the 10 studies was uploaded to the SyRF platform. I created a data extraction template - a list of questions on SyRF detailing the information to extract from each study. Data were extracted by two independent reviewers within the CAMARADES group (Myself and Emily Sena) and discrepancies resolved by discussion. This allowed us to determine the feasibility of the data extraction template we had designed, to develop training materials for reviewers, and to identify any potential issues prior to finalising our protocol. Furthermore, I identified an issue with the OFT regular expression in that it had not been entirely case-sensitive, hence including some spurious results by matching “oft” instead of “OFT”.

6.3.3 Study inclusion

We included studies which tested transgenic rodent models of AD in an open field paradigm, with no restriction on stage of development. We also included studies that investigated the effectiveness of treatments on these models. We defined the control population as a cohort of wild-type animals from the same litter as transgenic animals or an age-matched wild-type of the same background strain not subject to intervention. For treatment efficacy studies, we defined the control population as a transgenic AD model cohort which has not received any treatment intervention or has received appropriate sham treatment (e.g., injection of vehicle). Table 6.2 details our inclusion criteria in full.

	Inclusion criteria	Exclusion criteria
Type of study design	All primary experiments in transgenic AD animal models	Studies without a proper control and non-primary research including reviews, commentaries, and editorials.
Type of animal model	Transgenic AD animal models 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.
Type of intervention	All primary experiments in transgenic AD animal models, including studies which test the effect of a pharmacological or behavioural treatment given *in vivo.*	Studies with no controlled *in vivo* data (e.g. transgenic model + treatment vs transgenic model + vehicle).
Outcome measures	Studies which use the open field paradigm to assess locomotor activity, anxiety like behaviour, or exploratory behaviour	Studies which do not utilise the open field paradigm or only use the open field box for other behavioural tests e.g. novel object recognition
Language	All languages	None
Publication date	All publication dates	None

Table 6.2: Inclusion criteria for OFT SR

6.3.4 Research question

In this review, I aimed to investigate the following questions:

1. What experimental design variables are reported in studies using the OFT and what is the influence of these factors on locomotor activity and anxiety measurements?

2. How do different transgenic AD models perform on the OFT and what is the impact of treatments on OFT outcomes?

6.3.5 Study characteristics

The characteristics extracted from included publications are shown in Table 6.3.

	Information extracted	Additional details
Study meta-data		
	Title	Obtained from Endnote
	DOI	Obtained from Endnote
	First author	Obtained from Endnote
	Corresponding author	Obtained from Endnote
	Year	Obtained from Endnote
	Journal name	Obtained from Endnote
	Country of origin of corresponding author	
Animal husbandry		
	The light cycle	Light cycle hours (e.g. 12) and whether this was a reversed cycle (e.g. light in evening)
	The number of animals per cage	
	Presence of environmental enrichment	E.g. nesting material or toys in cages
Model induction		
	The animal species	
	The animal background strain	
	The transgene/ genetic manipulation	
	The sex of the animals	
	The source of animals	
	The type of model control	E.g. wild-type littermates or same background strain
Outcome information		
	The age of the animals at time of outcome assessment	Where reported as a range, we calculated the mean
Open field test protocols		
	Length of test arena (cm)	
	Width of test arena (cm)	
	Height of test arena (cm)	
	Light intensity of test arena (lux)	

	Length of time OFT recorded (minutes)	
	Number of OFT trials	
	Time habituated to test arena (minutes)	
	Method of automation	Including specific software where reported
	Test arena wall colour	
	Test arena shape	
Treatment information		
	Drug dose	
	Drug dose units	
	Number of times drug was given	
	Route of delivery	
	The length of drug treatment	
	The age of the animals at time of treatment	Where reported as a range, we calculated the mean
	Time between the administration of the treatment and outcome measurement	
	Details of anaesthesia (if used for treatment)	
Risk of bias and methodological quality		
	Reporting of random allocation of animals to treatment/control groups	
	Reporting of blinded assessment of outcome	
	Reporting of animal/data exclusions	
	Reporting of a sample size calculation	
	Reporting of approval by animal welfare committee	
	Reporting of a potential conflict of interest	
	Whether a study protocol is available dated before the experiments began	

Table 6.3: Information extracted from OFT publications

6.3.6 Methods for crowdsourcing data extraction

I recruited a crowd of reviewers from within the EQIPD consortium and invited external collaborators to contribute. Five training papers were selected and two

independent reviewers within our group annotated each paper. Disagreements were discussed to determine “gold standard” annotations. I created a detailed annotation guide (Appendix C6.3) detailing each of the annotation questions on the SyRF project. To complete the training, reviewers read the guidance materials and independently extracted information from each of the training papers on SyRF. Following this, I reviewed their annotations by running a custom R code to format the SyRF annotations in a readable format and comparing to the gold-standard annotations. I provided feedback to each reviewer individually to address any mistakes they had made.

Several webinars were delivered to introduce the process, and one to one video sessions were also used to answer detailed questions about methodology. I added reviewers to a Microsoft Teams channel where I collated the guidance resources and encouraged questions via the chat functionality. Throughout the process, I wrote a custom R code to summarise up to date progress on the project and sent frequent updates in the form of online web pages generated in R (Figure 6.1). The link to the latest report was distributed to reviewers via email and Teams on a regular basis.

EQIPD OFT Review Progress Report

Kaitlyn Hair

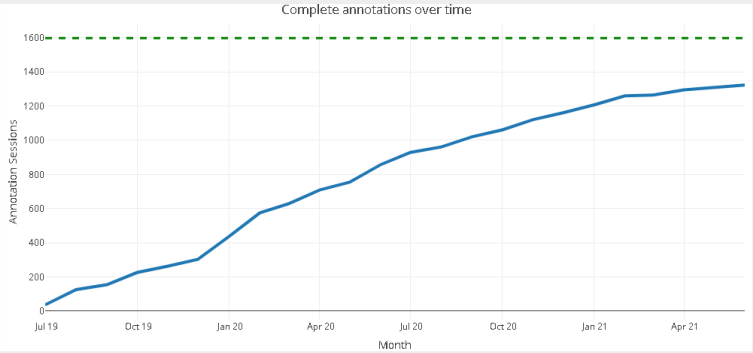
30 June, 2021

Reviewers working on the project

A total of **84** reviewers have registered on the SyRF platform since the beginning of the project. **33** have started the reviewer training and **26** reviewers have now completed the training.

Progress over time

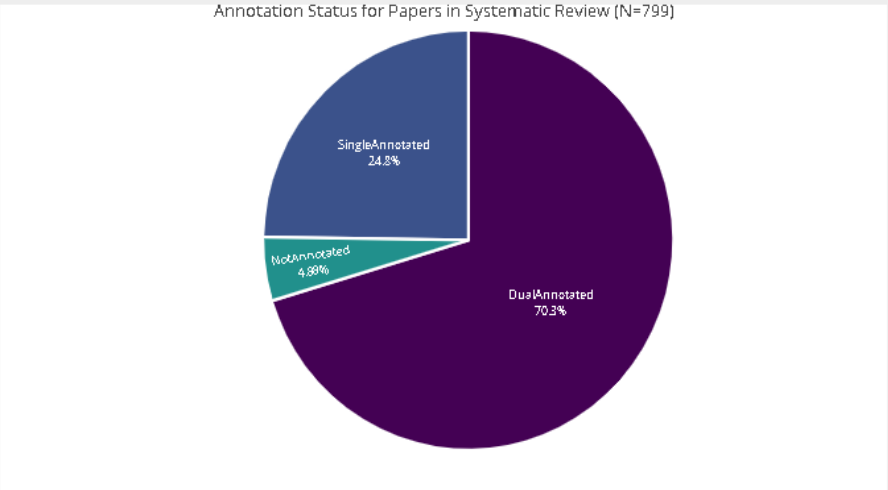
With a total of **799** to extract from, there are a total of **1598** annotation sessions required (2 reviewers for each paper). The number of annotation sessions over time is shown below. The green line indicates the total number of sessions required to complete the project. The project is **82.7%** complete!



Data extraction progress

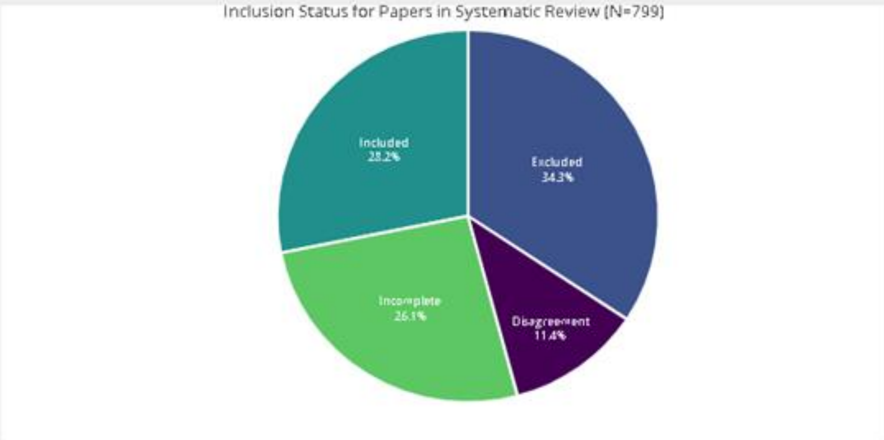
From a total of **799**, there are **39** papers which no reviewers have extracted from.

760 papers have been extracted by at least one reviewer and of these papers, **562** have been extracted by 2 independent reviewers and **198** have been extracted by only one reviewer.



Screening progress

Of the **562** papers that have been extracted by 2 reviewers, **214** papers have been included by both reviewers and **261** have been excluded from the review. Reviewers disagree on whether **87** papers should be included or not.



Reviewer of the Fortnight!

****

Our reviewer of the fortnight is **Louis Dwomoh** who completed 7 papers over the past 2 weeks. Amazing work!

****

Reviewer Leaderboard

Finally, here is our reviewer leaderboard. This leaderboard does not take into account the 5 training papers or papers that are incomplete (saved) papers.

Annotator	Papers_Completed
Isabel	186
Kaitlyn	162
Farah	121
Lukas	115
Bettina	102
Isabel S	62
Nadia	56
Aishwarya C	56
Vincent	54
Daniel	48
Louis	47

Figure 6.1: Exemplar progress update sent to reviewers

Publications identified using the OFT regular expression dictionary were uploaded to SyRF for data extraction. We extracted study characteristics and numerical outcome data (mean, SD or SEM, and sample size) from included publications and entered the data into a SyRF project. When outcome data were presented in a graphical format, we used Web Plot Digitizer (https://automeris.io/WebPlotDigitizer/) to measure means and errors bars. We aimed for each publication to be assessed by two independent reviewers and reconciled by a third independent reviewer. In this subset analysis I present only single annotated data. For any publications which omit vital information, reviewers recorded the publication information on a shared spreadsheet, so that authors could be contacted later. This process will be completed by the time the entire review is fully reconciled.

6.3.7 Methods for selecting APP modelling sub-set

To select a subset of OFT data for analysis, I wrote an R code to retrieve and format data from SyRF, clean the data, and select studies where the model was one of the commonly used APP models discussed in Chapter 1 (Table 1.1). I also summarised the data, generated visualisations, and performed the meta-analysis within R. Where two reviewers extracted data from a publication, the annotations from senior reviewers (who had reviewed a substantial number of publications) were preferentially retained. This code-based method was preferred over other approaches as it facilitated reproducibility and transparency. Where a simple error was spotted (e.g., a typo or an alternative description of an outcome or model), I corrected it within the analysis code. Where larger errors existed (e.g., omission of a cohort of animals or a missing outcome measure), where dual-data extraction had been completed, I preferentially selected the other reviewer's annotations. In a small number of cases, I re-extracted the data as an additional reviewer (if I had not been one of two original reviewers).

6.3.8 Random effects meta-analysis

Due to uncertainty around the true biological meaning of each OFT measure, and the extent to which each outcome represents anxiety and/or activity, I analysed each outcome with sufficient data (N ≥10 publications and N ≥25 comparisons) separately.

I calculated a standardised mean difference (SMD) for each individual modelling versus wild-type comparison for each outcome measure. Occasionally, very similar outcomes were measured using different units e.g., “time in centre as (% of overall time)” and “time in centre (duration)”. Further, due to open field configurations, the peripheral areas may be termed “corners”, “periphery”, or “corners”. In these instances, I grouped together the outcomes and calculated a nested effect size in the analysis according to previously described methods (Vesterinen et al., 2014).

Where a single wild-type control group served multiple transgenic intervention groups, this was adjusted for in the analysis by dividing the size of the control group by the number of groups it served. For repeated measures, I retained only the first time point measured to ensure that the data were comparable. For studies where measurements were only taken in multiple bins (typically of 1-5 minutes duration) over one OFT trial, I calculated the area under the curve. Following methods described previously for preclinical meta-analysis (Vesterinen et al., 2014), Hedge’s G effect sizes were calculated to obtain a standardised mean difference (SMD) for each modelling comparison (transgenic group versus control group). Effect sizes were weighted based on the standard error of each study, with more precise studies given greater weight in the meta-analysis. I pooled SMD effect sizes for each outcome using a random-effects model with a restricted maximum likelihood (REML) estimate of between study variance to get an overall effect size.

6.3.9 Multivariable meta-regression

I conducted a multivariable meta-regression to identify sources of heterogeneity. Although univariable analysis is the conventional approach in pre-clinical meta-analysis, a multivariable method was preferred as it has been shown to explain a greater proportions of the heterogeneity (Tanriver-Ayder, Faes, van de Casteele, McCann, & Macleod, 2021). To explore the heterogeneity in the dataset, I performed a model building validation exercise to fit a multivariable meta-regression model following previously described methodology (Harrer, Cuijpers, Furukawa, & Ebert, 2021; Tanriver-Ayder et al., 2021). I began with performing univariable meta-regression to determine the proportion of heterogeneity accounted for by each variable. The most significant variable was included in the model, and then other covariates were added separately in combination with the first variable. To determine if another covariate should be added, I recorded the change in Akaike's information criteria (AIC) by running a likelihood ratio test to compare models. Where the new model (with the covariate) lowered the AIC and the likelihood test was significant, this indicates an improvement in the model i.e., it explains a greater proportion of the heterogeneity. For each interaction, I added the covariate which lowered the AIC by the greatest amount and then repeated the process by adding another covariate. The model was considered complete when the AIC showed no further reductions on the addition of new covariates, or any additional variables did not significantly improve model fit.

Variables which were reported in less than 25 experiments were not included as a covariate in the multivariable meta-regression model. Furthermore, if a numerical continuous variable was missing in more than 10% of experiments, it was not included as a covariate, as this was deemed to remove too much useful data from the analysis. For categorical variables, if a category applied to less than 10 experiments, it was grouped into an "Other" category or combined, where

reasonable, with another category. For categorical variables with less than 10 in one category and only two meaningful categories e.g., “Reported” and “Not reported”, that variable was not included in the heterogeneity analysis.

To reduce multi-collinearity, two variables (height and width of the OFT arena) were combined into one variable for the analysis. Area was calculated depending on OFT shape (circular or rectangular). Due to poor reporting of habituation time, I made this variable categorical (Habituation reported, no habituation reported), with the reasoning that any habituation may have had some impact compared to no habituation.

6.4 Results

6.4.1 Crowd recruitment, training, and collaboration

Eighty-nine reviewers signed up to the project on SyRF. Of these, 45 began the reviewer training, with 33 reviewers going on to complete it. Seventeen reviewers annotated 35 or more publications (the pre-specified criteria for co-authorship).

6.4.2 Identification of relevant publications

The open field test regular expression identified 799 potentially relevant publications within the preclinical AD dataset. A small number of additional duplicates were identified during data extraction or analysis and removed from further evaluation, leaving 781 publications in the dataset.

We identified 237 publications describing modelling and intervention experiments in selected APP transgenic models where behaviour was evaluated in the OFT. Extracted data were split into two datasets – (i) a modelling dataset comparing APP transgenic animal models groups to an appropriate wild-type control (n=192), and (ii) a treatment dataset comparing treated APP transgenic model groups to APP

transgenic controls (n=133). For modelling experiments, I included experiments where a vehicle or sham treatment had been given to both wild-type and transgenic animal groups. A PRISMA flow diagram of included studies for the sub-set is shown in Figure 6.2.

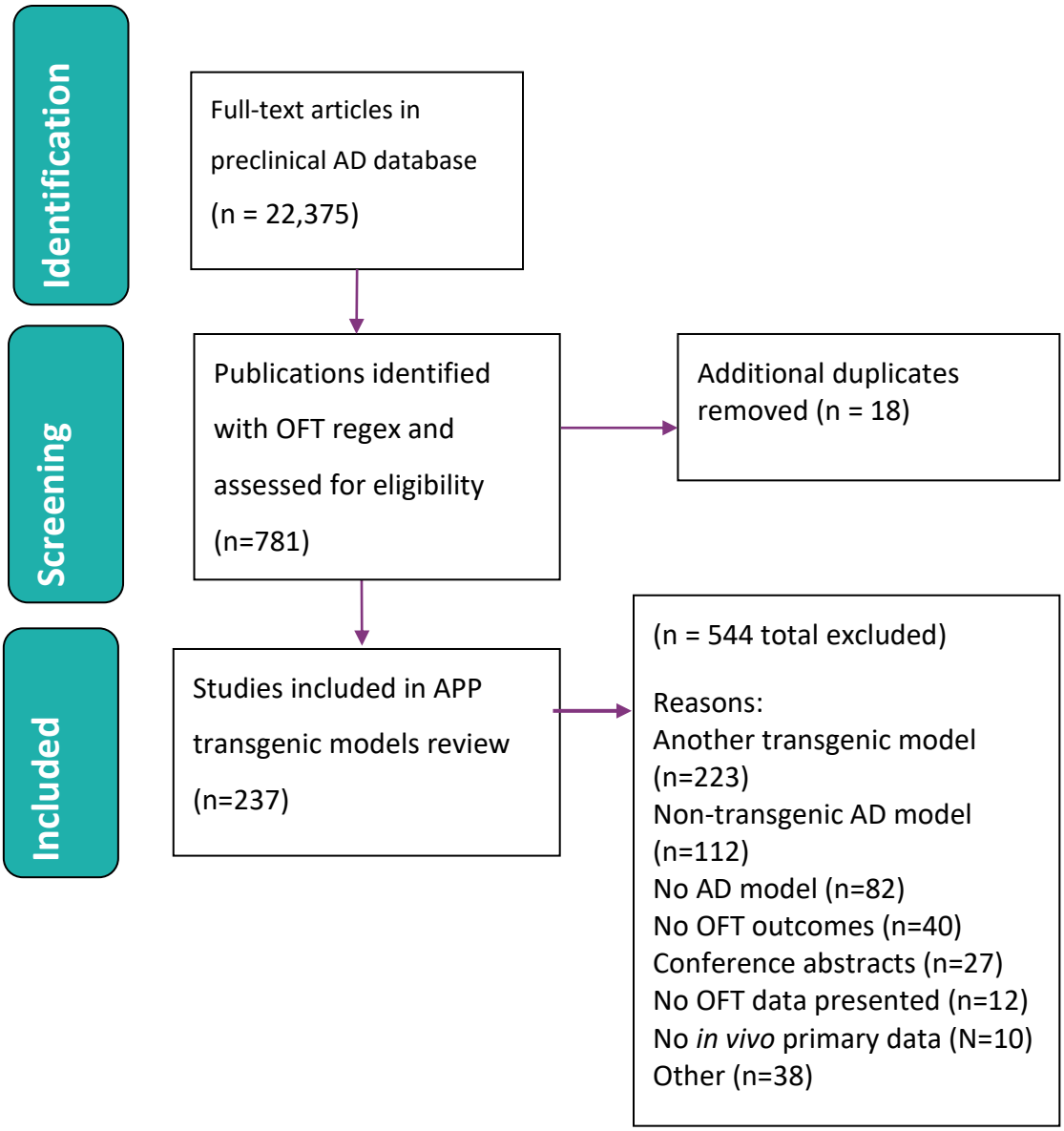


Figure 6.2: PRISMA flow diagram
Note: many papers will be excluded for several reasons. The order of exclusions indicates the main reasons, where a paper will be excluded for not being in a non-transgenic model before not measuring OFT outcomes

6.4.3 Publications over time

Since the first relevant publication in 1999, the number of papers has increased over time (Figure 6.3). Between 2008-2017, there has been an average of 20 new papers published each year in relevant APP models. It should be noted that the number of papers from 2018 is inappropriately small due to the timing of the search.

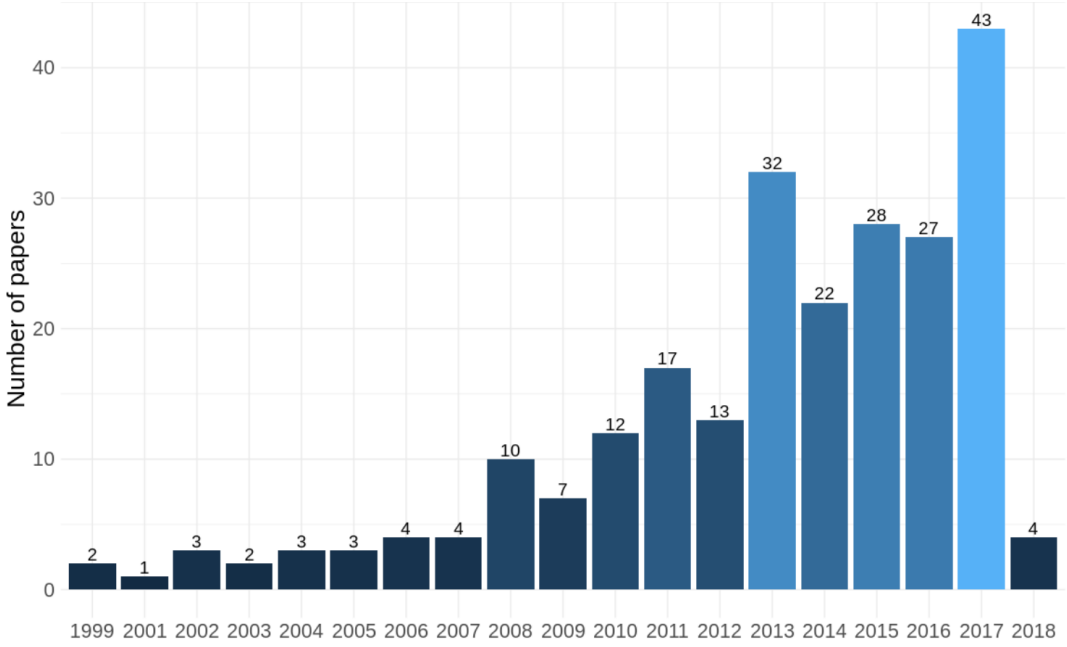


Figure 6.3: OFT publications per year in common APP models

6.4.4 Research location

For the majority of included publications, the corresponding author was based in the United States (N=89). Many publications were also identified from China (N=31), Germany (N=26), Spain (N=22), and Canada (N=14). Figure 6.4 shows the country distribution of included publications. This may not be a true representation of where most relevant research is being carried out. Instead, it may simply reflect the

countries where most research is indexed within the databases I chose to search.

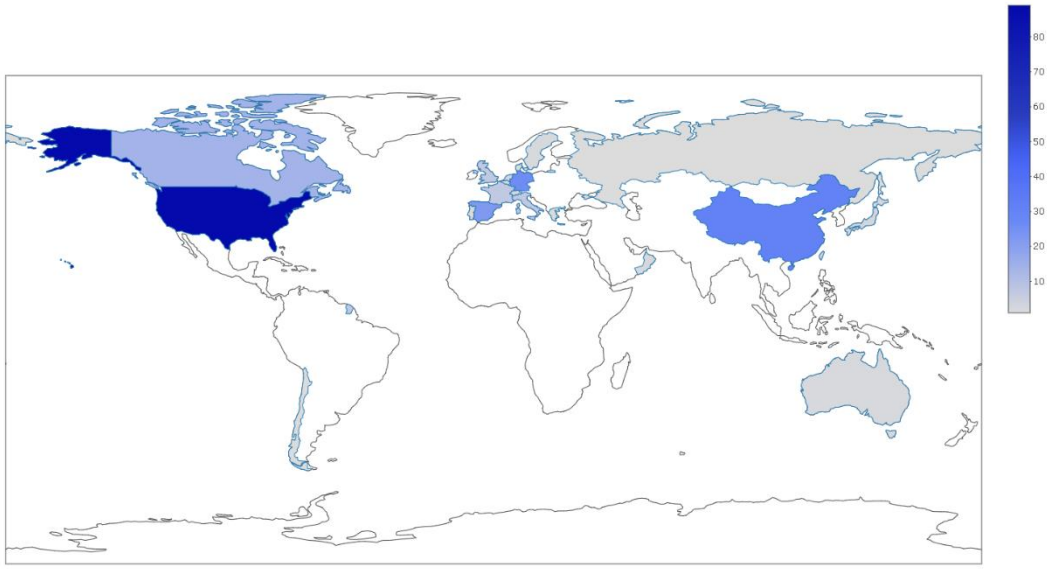


Figure 6.4: Country of corresponding author for included publications

6.4.5 Transgenic model characteristics

A summary of the number of publications with modelling and treatment experiments using each transgenic model is listed in Table 6.4. The APPSwe/PSENEN1de9 model was most frequently used in the modelling dataset (n=38) and treatments were most frequently tested in the 3xTg-AD model. Overall, for the modelling dataset, there were n=68 publications describing experiments in transgenic APP models, n=92 publications in APP/PS1 models, and n=34 in APP/PS1/MAPT models. In the treatment dataset, there were n=46 experiments in APP models, n=61 in APP/PS1 and n=26 in APP/PS1/MAPT models. A full list of publications, models used, and sex of the animals is shown in Appendix 5.1. The background strain of the transgenic model was not reported in 70/237 publications (29.5%) and the background strain of the control was omitted in 66/237 publications.

Model	Mutations	Modelling dataset	Treatment dataset
APPSweLon	APP	1	0
PDAPP	APP	2	0
Tg-SwDI	APP	2	6
TASD41	APP	4	0
PS/APP	APP/PS1	4	6
APP23	APP	7	3
APPSwe/PSEN1(A246E)	APP/PS1	7	4
TgCRND8	APP	13	12
J20	APP	14	7
APPPS1	APP/PS1	19	11
5xFAD	APP/PS1	24	14
Tg2576	APP	25	18
3xTg-AD	APP/PS1/MAPT	34	26
APPSwe/PSEN1dE9	APP/PS1	38	26

Table 6.4: Summary of models used across publications

6.4.6 Interventions

A total of 133 publications in the dataset tested the effect of a treatment intended to improve AD pathology. Only 4 treatments (Exercise, Levttiracetam, Melatonin, Paroxetine) were reported in more than 1 publication (Figure 6.5). We decided not to perform a meta-analysis of this dataset due to the low sample size and variation in transgenic models on which these treatments were tested. Treatments for each publication are listed with transgenic models and sex of animals in Appendix 5.1.

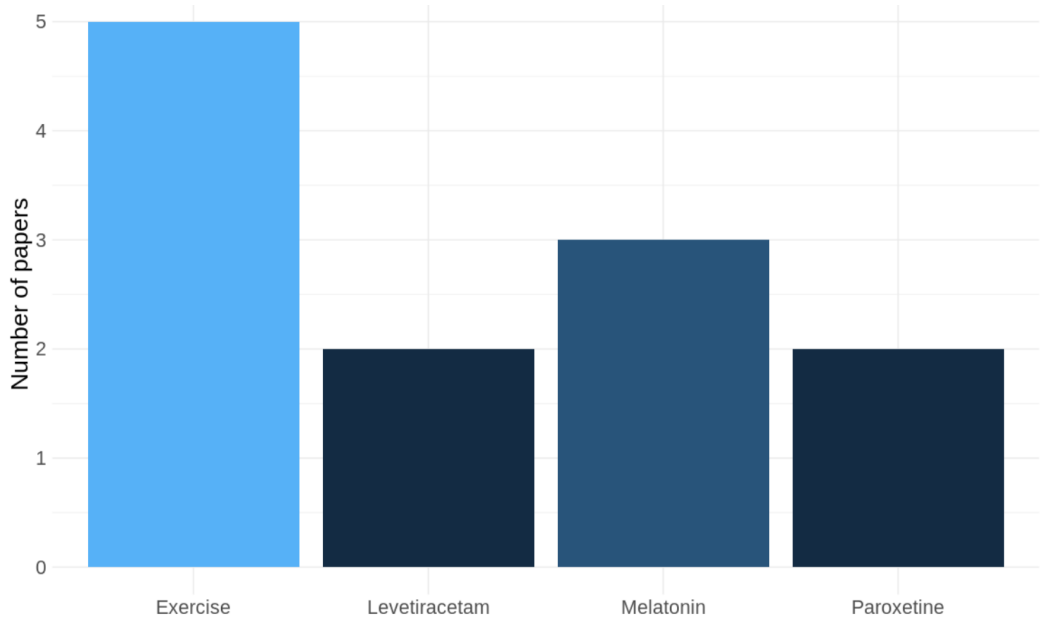


Figure 6.5: Summary of interventions used in included publications
Only interventions reported in more than one publication are shown.

6.4.7 Outcome measures

Forty-eight different OFT behavioural measures were reported across modelling experiments (Table 6.5). The most reported measure in both the modelling and treatment datasets was total distance travelled (N=127 publications and N=90 publications respectively).

Outcome measured	Modelling dataset (N=192)	Treatment dataset (N=133)
Activity		
Total distance travelled	127	90
Line / grid crossings	35	22
Velocity	19	20
Inactivity time	11	10
Total activity count	7	7

Outcome measured	Modelling dataset (N=192)	Treatment dataset (N=133)
Time spent moving	5	4
Latency to move	5	2
Inactivity counts	4	3
Fast-moving time	1	1
Fast-moving time in centre	1	0
Fast-moving time in periphery	1	0
Highly mobile horizontal activity counts	1	1
Inactivity time in centre	1	0
Inactivity time in periphery	1	0
Maximum velocity	1	1
Slow-moving time	1	1
Slow-moving time in centre	1	0
Slow-moving time in periphery	1	0
Velocity in centre	1	0
Defecation		
Defecation frequency	11	7
Grooming		
Grooming time	8	4
Grooming episodes	6	7
Latency to groom	3	2
Grooming (no further detail)	2	1
Urination		

Outcome measured	Modelling dataset (N=192)	Treatment dataset (N=133)
Urination	3	2
Rearing		
Number of rears	41	23
Latency to rear	4	1
Rearing (vertical time)	5	3
Rearing (no further detail provided)	1	1
Rears against wall	1	0
Number of rears in centre	0	1
Number of rears in periphery	0	1
Thigmotaxis		
Time in centre	72	56
Time in periphery	27	29
Distance in centre	23	13
Centre entries/ crossings	14	7
Periphery entries/crossings	8	3
Distance in periphery	8	5
Latency to exit centre	3	4
Latency to periphery	2	2
Activity in centre	1	0
Latency to enter centre	1	0
Latency to exit periphrery	1	0
Thigmotaxis (distance ratio)	0	1

Outcome measured	Modelling dataset (N=192)	Treatment dataset (N=133)
Wall contact duration	1	1
Habituation		
Change in movement in later trial	2	2
Other		
Stereotypy (no further detail)	1	1
Stretching frequency	0	1

Table 6.5: Summary of outcomes reported across publications

6.4.8 Age of animals

The median age of mice given a treatment intervention was 5.0 months (interquartile range: 3.0 – 8.3 months). Outcomes were measured at a median age of 8.9 months (6.0 – 12.3 months interquartile range).

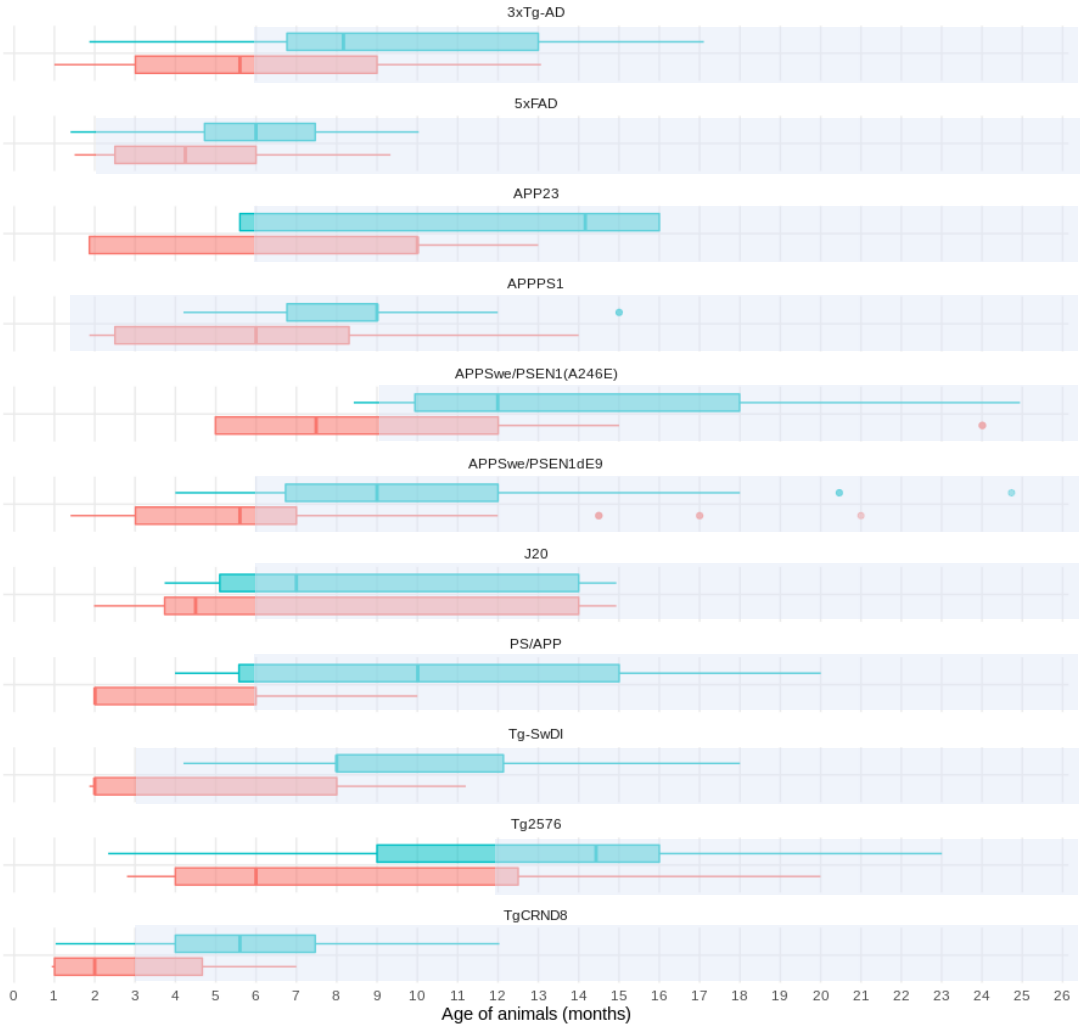


Figure 6.6: Boxplot of distribution of age of animals at time of intervention and outcome assessment.
The distribution of age at which OFT outcomes were measured (blue) and age at which treatment was administered (red) are shown for each APP transgenic model. The lilac shaded area indicates when amyloid pathology is present (source for age of amyloid pathology: Alzforum website).

6.4.9 Sex of animals

Overall, the sex of animals used was not reported for at least one comparison in 8% of publications (19/237). Mixed groups were exclusively used in 51/237 publications and 14 publications analysed male and female mice cohorts separately. 39.2% of publications (93/237) measured outcomes on male mice only, compared to 24.9% (25/237) of publications which tested only female mice. The sex of animals reported in each publication is shown in Appendix C6.5.

6.4.10 Study quality and risk of bias

Across the included publications, many failed to report measures to reduce the risk of bias. Blinded outcome assessment was reported in just under a quarter of publications 24.4% and exclusion criteria were reported in 15.5% of publications. Power calculations to determine an adequate sample size were only mentioned in 5/237 (2.1%) of publications. Randomisation and allocation concealment are not relevant for transgenic modelling experiments, as the intervention is randomly pre-determined, as the mutated genes are passed on to offspring in a mendelian fashion. In applicable publications with treatment interventions, 29.7% reported randomisation and only 6.6% reported that allocation of treatment groups had been concealed. Measures of study quality were more frequently reported. One hundred and thirty-six publications (57.6 %) had conflict of interest statements and 70.9% reported ethical approval by an institutional animal welfare committee. The reporting of relevant risk of bias and study quality measures is summarised in Table 6.6. We also recorded if any publications had pre-registered a protocol prior to the study, but no publications reported this. RoB reporting is shown by publication in Appendix C6.7.

	N reporting	N total	Percentage reporting
Blinding	57	237	24.1 %
Randomisation	38	136	27.9
Allocation concealment	9	136	6.6 %
Ethical approval	168	237	70.9%
Conflict of interest statement	136	237	57.4 %
Exclusions	36	237	15.2 %
Power calculation	5	237	2.1 %

Table 6.6: Reporting of study quality measures and reporting of measures to reduce the risk of bias.
N indicates the number of publications.

6.4.11 Animal husbandry

Most publications (68.4%) reported the light cycle the animals were maintained in. Of those reporting a light cycle, 155/162 used a standard, 12-hour lights on-off cycle in the animal housing area, where lights are on during the day. Two publications reported a reversed 12-hour cycle (dark during the day), and two publications reported prolonged 13-hour and 14-hour light cycles. 39.2% of publications reported the number of animals per cage. Of these publications, 29 reported that mice were housed individually, and others reported a group of 2-15 mice per cage. Environmental enrichment was rarely reported, with just 20/237 publications (8.4%) mentioning additional materials or stimulation to mimic a rodent's natural environment and/or reduce stress. The reporting of husbandry details is summarised in Table 6.7

	N reporting	N total	Percentage reporting
Environmental enrichment	20	237	8.4 %
Light cycle	162	237	68.4 %
Number of animals per cage	93	237	39.2 %

Figure 6.7: Reporting of animal husbandry details
N indicates the number of publications.

6.4.12 Open field protocols

There was substantial variation in the reporting of open field parameters. The dimensions of the OFT were largely reported, with length and width reported in 86.9% (206/237) of publications. The height of the open field wall was less frequently reported, with details reported in 63.7% of studies. Time habituated was only reported in 10.5% (25/237) publications. Light intensity measured by luminance flux per unit area (lux) in the open field was only reported in 11.8% (28/237) publications. One additional paper reported wattage, which cannot be converted into light intensity without further information. The duration OFT behaviour was recorded for was reported 93.2% of publications (221/237). It was often unclear if animals were tested in the OFT in only one trial, or whether the results of multiple trials were averaged. The number of trials was clear to reviewers in 130/237 publications. Histograms showing the distributions of OFT protocols is shown in Figure 6.8.

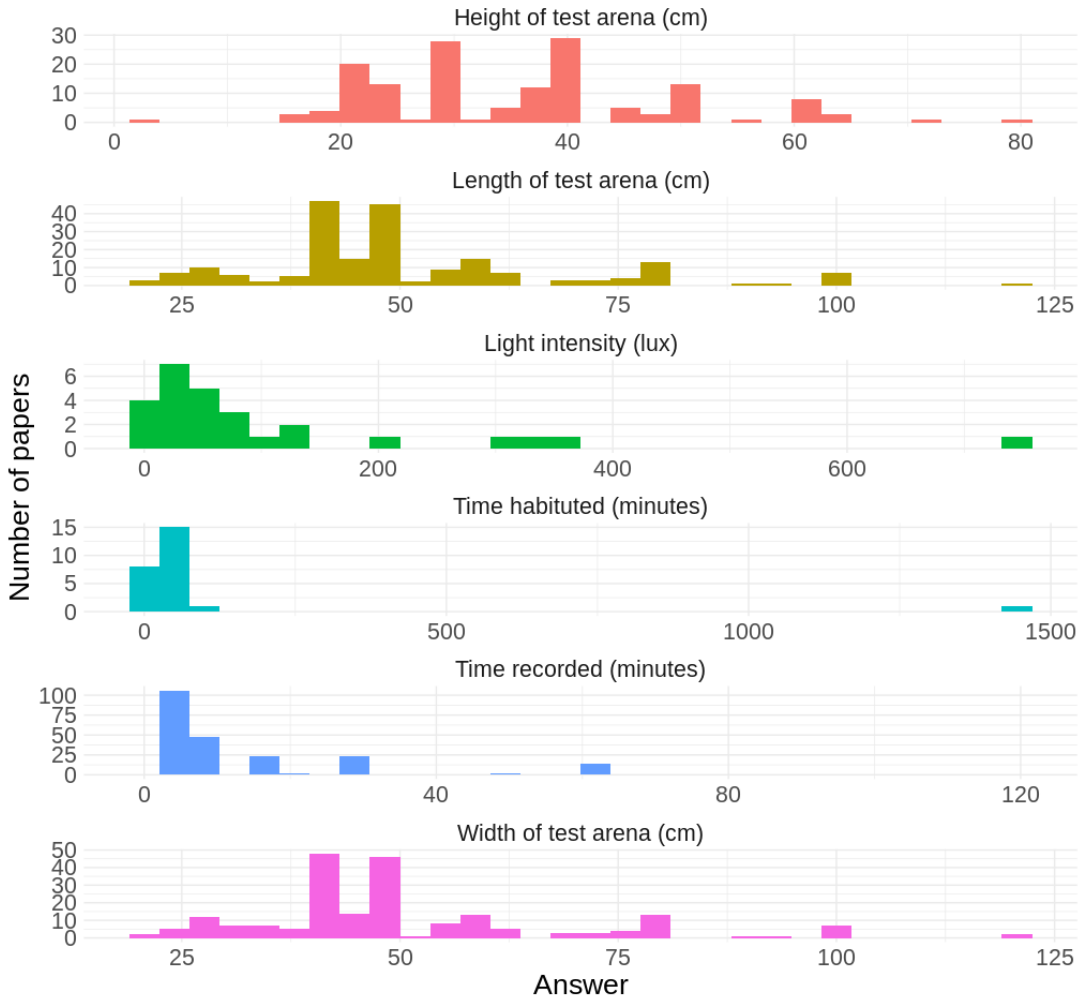


Figure 6.8: Histograms representing the reporting of OFT parameters across publications. Individual histograms only include publications which report each parameter.

The shape of the OFT arena was reported in 81% (192/237) of publications and wall colour was reported in 99/237 (41.8%). The most common shape for the OFT arena was rectangular/square (177/237). The most common wall colour was white (43/237). To illustrate the variety of open field arenas used in modelling experiments, I created waffle chart (i.e., square pie chart), with different icons representing the proportion of each colour and shape combination reported within the literature (Figure 6.9). Rectangular/square open field arenas with an unknown

wall colour were the most frequently reported. A full list of all OFT protocols reported in each included publication are shown in Appendix C6.6.

Figure 6.9: Proportion of publications using different OFT shapes and wall colours
The number of icons is proportional to the number of publications. NR = Not reported. Red indicates an open field arena with an unknown wall colour. Question marks indicate unknown open field shapes.

6.4.13 Meta-analysis of transgenic modelling interventions

From 192 papers reporting modelling experiments, only 175 had comparison data which could be included in the meta-analysis. The main reasons for exclusion from the meta-analysis were the use of median instead of mean (n=4), missing or unreadable SEM or SD measurements (n=10) and missing sample sizes (n=14). Further, there was evidence of selective outcome reporting, where some papers mentioned several OFT outcomes in the methods, but only reported numerical data for a few outcomes (e.g., where a significant difference was identified between the transgenic and control group).

There were over 25 experimental comparisons for six outcome measures: total distance travelled, time spent in the centre, number of rears, number of grid/line crossings, time in periphery, and distance in centre. The number of publications and independent comparisons for each outcome sub-set is shown in Table 6.7.

Outcome	N papers	N comparisons
Total distance travelled	105	150
Time in Centre	58	75
Number of rears	37	61
Line / grid crossings	33	58
Time in Periphery	25	30
Distance in Centre	22	29

Table 6.7: Summary of publications and experimental comparisons for each OFT outcome included in the meta-analysis

Variables included in the univariable analysis were:

- **Animal model details** (transgenic model category, transgenic model, background strain of transgenic model cohort, sex of animals, age of animals at time of outcome measurement)
- **Open field test protocols** (height of arena, area of arena (calculated value from width and length), duration of OFT, arena wall colour, and arena shape)
- **Relevant risk of bias and study quality items** (reporting of blinded outcome assessment, reporting of exclusions, reporting of animal welfare committee approval, and conflict of interest statement)

There were missing data for several continuous variables: age, light intensity, OFT duration, arena height, and arena area. Reporting of a power calculation, and pre-registration of a protocol were not included as variables in the analysis due to <10 publications reporting these measures. Light intensity was not included in any

analysis as it too few comparisons reported this (n<25). Height and area of the OFT arena were not considered in any of the multi-variable meta-regression models as there were data missing in over 10% of comparisons. Habituation (as a binary variable reported/not reported) was only reported in >10 publications for total distance travelled and was not analysed for any other outcome.

1. ***Total distance travelled***

In modelling experiments measuring total distance travelled, the pooled effect of modelling interventions was 0.12 SMD (95% CI -0.09 to 0.34, n=150 comparisons), meaning that transgenic animals travelled a greater distance in the open-field versus wild-type controls. Heterogeneity was high (I^2 = 96.1%, Q=3785.75, Tau^2=1.65). A forest plot of comparisons is shown in Figure 6.10.

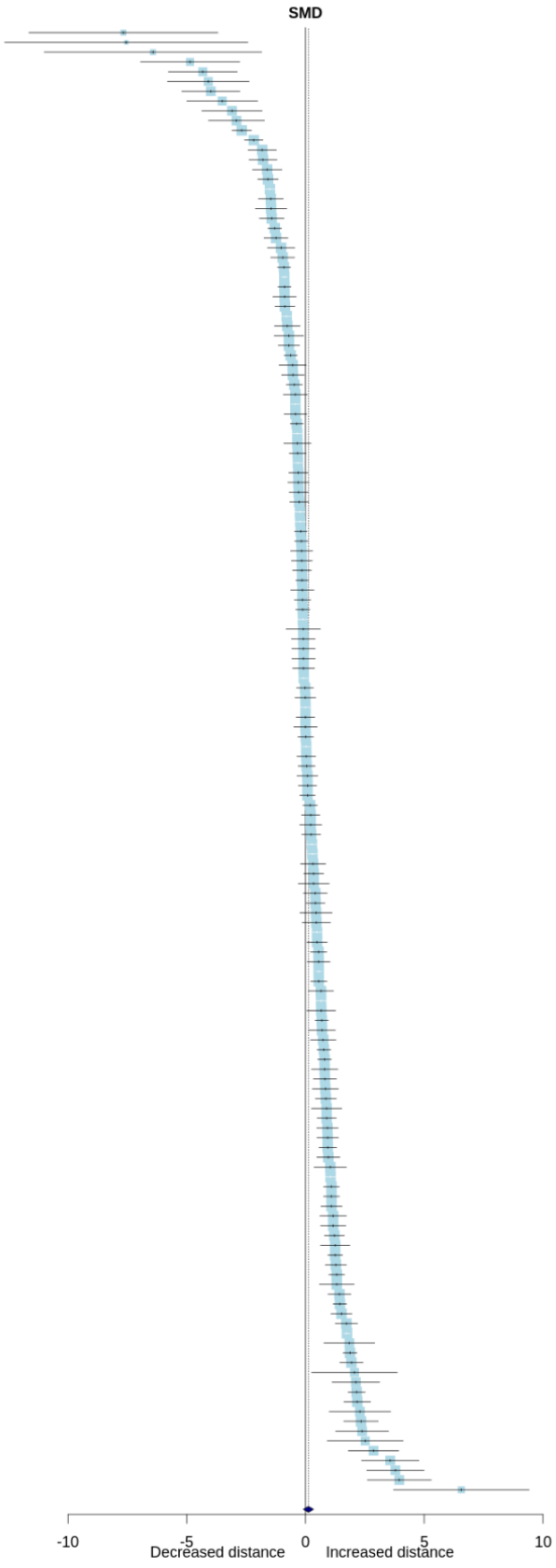


Figure 6.10: Forest plot of effect sizes for total distance travelled modelling comparison.
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. A high resolution size image with more detail is available at: https://osf.io/va745/

In the univariable regression analysis, transgenic mouse model and model category each explained a significant proportion of the heterogeneity (Tables 6.8-6.9). TgCRND8 and Tg2576 mice travelled a greater distance in the open field, while 3xTg-AD mice travelled less. A similar pattern is seen in the transgenic model category univariable analysis, where APP transgenic models had a higher estimate of total distance travelled and APP/PS1/MAPT models had a lower estimate. Reporting of blinded outcome assessment was associated with a higher estimate of effect in the univariable analysis. In the multivariable meta-regression, I began with transgenic model as a predictor. No additional variables significantly improved the model fit.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
3xTg-AD	-0.86	-1.41 to -0.31	<0.0002	1.32	97.44	19.91
5xFAD	-0.41	-1.18 to 0.36				
APPPS1	0.24	-0.58 to 1.07				
APPSwe/PSEN1dE9	0.12	-0.61 to 0.85				
Other	0.73	-0.09 to 1.55				
Tg2576	0.69	-0.08 to 1.46				
TgCRND8	1.02	0.02 to 2.02				

Table 6.8: Univariable meta-regression of total distance travelled with transgenic model N=150 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APP	0.71	0.33 to 1.09	<0.0001	1.37	97.6	17.02
APP/PS1	0.04	-0.45 to 0.53				
APP/PS1/MAPT	-0.86	-1.54 to -0.2				

Table 6.9:Univariable meta-regression of total distance travelled with model category N=150 comparisons, CI =confidence intervals.

*2. **Time in centre area***

In experiments measuring time in centre area, the pooled effect of modelling interventions was -0.12 (95% CI -0.36 to 0.12, n=75 comparisons) indicating that the time spent in the centre area was lower in transgenic models versus wild type.

Heterogeneity was high (I^2 = 96.4%, Q=2052.84, Tau^2=1.07). A forest plot of comparisons is shown in Figure 6.11.

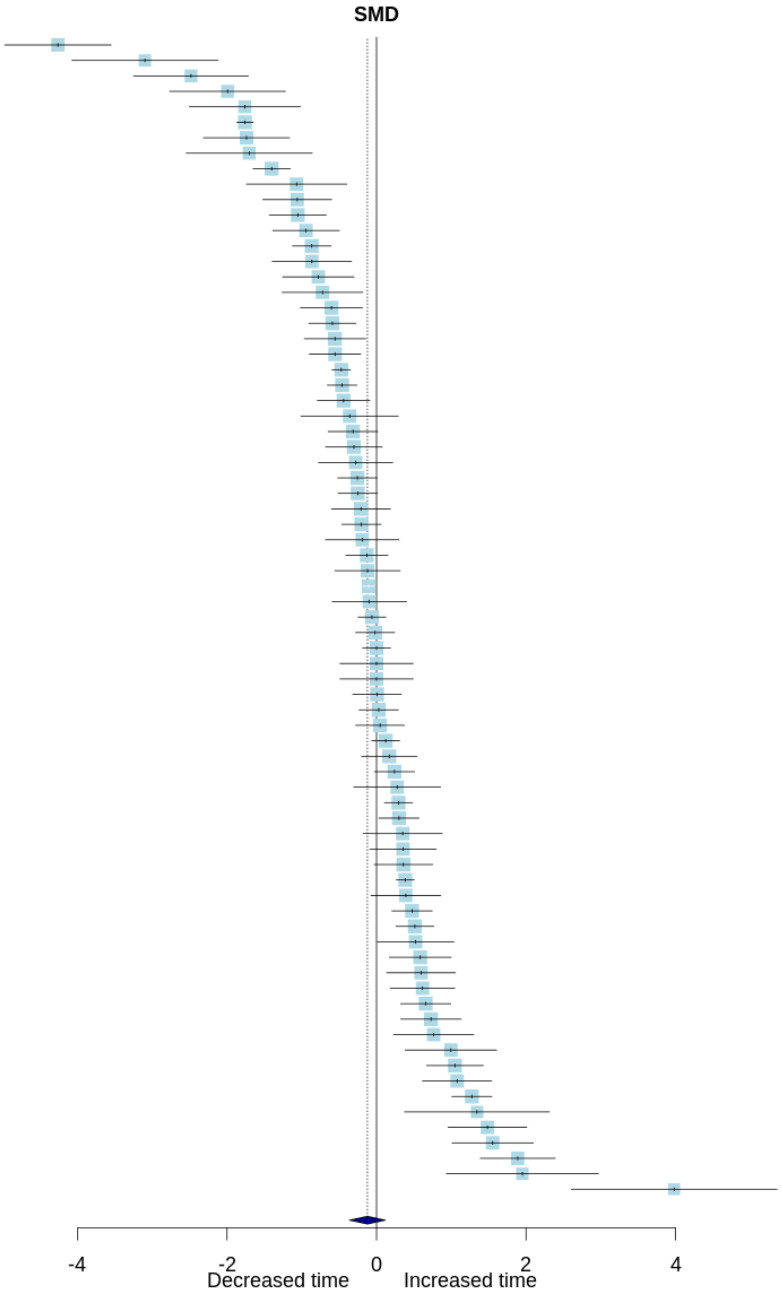


Figure 6.11: Forest plot of effect sizes for time in centre area.
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. A high resolution size image with more detail is available at: https://osf.io/sbehd/

Open field arena wall colour (Table 6.10) and strain (Table 6.11) each explained a significant proportion of the heterogeneity in the univariable analysis. Experiments in OFT arenas with white walls had a lower SMD estimate (reduced time in centre area), while experiments in arenas with transparent walls had higher SMD estimates. The impact of strain is difficult to assess as many studies omitted the background strain. According to the univeriate regression model, experiments in C57BL/6 mice were associated with higher estimates of effect. In the model building exercise, starting with the most significant predictor (wall colour), the addition of background strain led to a significant reduction in the AIC. No further variables improved the model. Therefore, the best model explained 15.54% of the heterogeneity (Table 6.12).

Variable	SMD	95% CI	P value	Tau^2	I^2	R^2
Wall colour: NR	0.17	-0.16 to 0.49	0.0013	0.95	97.69	11.5
Wall colour: Other	-0.3	-1.21 to 0.61				
Wall colour: Transparent	0.03	-0.7 to 0.76				
Wall colour: White	-0.82	-1.41 to -0.23				

Table 6.10: Univariable meta-regression of time in centre area with wall colour N=75 comparisons, CI =confidence intervals.

Variable	SMD	95% CI	P value	Tau^2	I^2	R^2
Background strain: C57BL/6	0.21	-0.18 to 0.6	0.03	1	97.71	6.96
Background strain: NR	-0.53	-1.07 to 0.02				
Background strain: Other	0.01	-0.65 to 0.68				

Table 6.11: Univariable meta-regression of time in centre area with background strain N=75 comparisons, CI =confidence intervals.

Variable	Regression weight (β coefficient)	95% CI	P value	Tau^2	I^2	R^2
Intercept	0.36	-0.05 to 0.76	0.007	0.9	97.42	15.54
Wall colour: Other	-0.33	-1.24 to 0.58		0.9	97.42	15.54
Wall colour: Transparent	0.29	-0.52 to 1.1		0.9	97.42	15.54
Wall colour: White	-0.77	-1.38 to -0.16		0.9	97.42	15.54
Background strain: NR	-0.69	-1.28 to -0.1		0.9	97.42	15.54
Background strain: Other	-0.19	-0.84 to 0.45		0.9	97.42	15.54

Table 6.12: Multivariable meta-regression of time in centre area with background strain N=75 comparisons, CI =confidence intervals.

3. ***Number of rears***

Transgenic animal models displayed rearing behaviour a lower number of times in the open field compared to wild-type controls (-0.84 SMD, 95% CI -1.18 to -0.50, n=61 comparisons). Heterogeneity in the dataset was high (I^2 = 93.9%, Q=978.9, Tau^2=1.68). A forest plot of comparisons is shown in Figure 6.12.

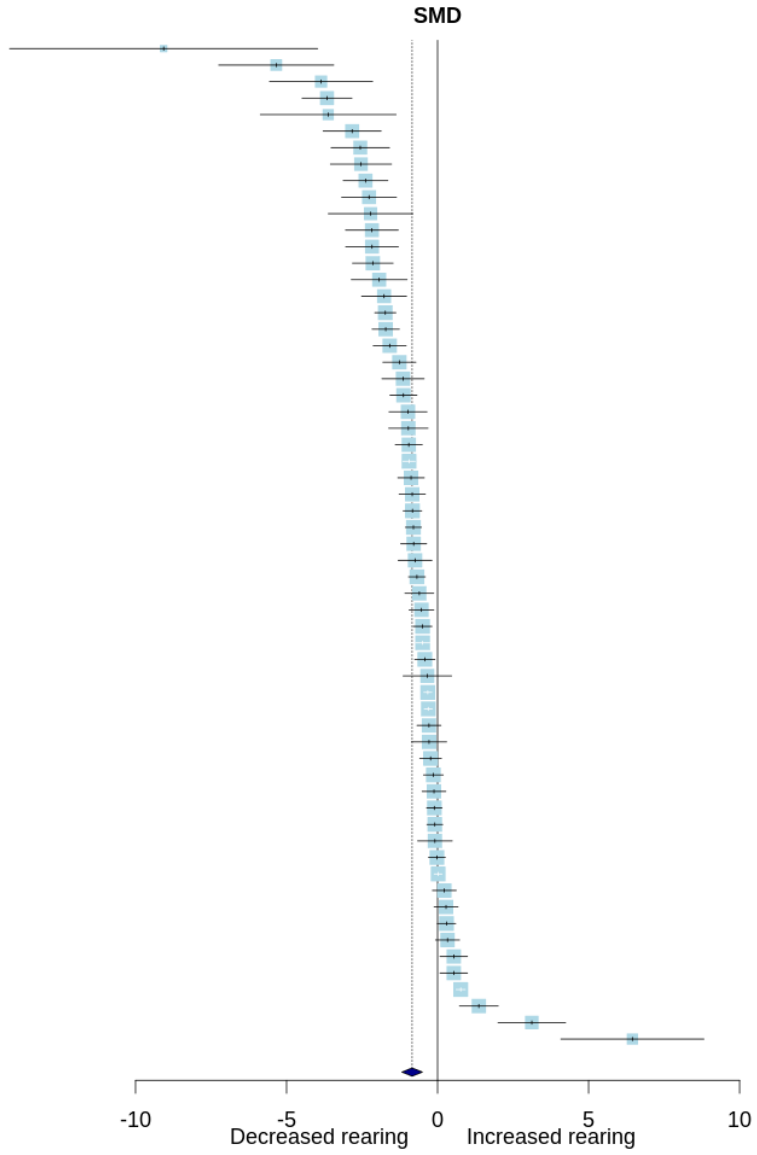


Figure 6.12: Forest plot of effect sizes for number of rears.

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. A high resolution size image with more detail is available at: https://osf.io/dcme6/

Transgenic animal model, model category, and height of the OFT arena each explained a significant proportion of the heterogeneity in the univariable analysis (Table 6.13 – 6.15). Experiments in APP/PS1/MAPT mice had lower SMD estimates (reduced rearing) in the open field. Similarly, in the univariable analysis of individual models, 3xTg (APP/PS1/MAPT) had lower SMD estimates than 5xFAD (APP/PS1) or other transgenic mouse models. Greater arena height was associated with increased rearing behaviour (Figure 6.13). However, height could not be taken forward into the multivariable model as too much data were missing (43/61 comparisons reported wall colour). In the model building exercise, I began with transgenic model as a single predictor. No variables significantly improved the initial model.

Variable	SMD	95% CI	P value	Tau^2	I^2	R^2
Model: 3xTg-AD	-1.77	-2.37 to -1.17	0.0001	1.01	96.81	40.06
Model: 5xFAD	-1.41	-2.42 to -0.39				
Model: Other	-0.15	-0.89 to 0.59				

Table 6.13: Univariable meta-regression analysis of number of rears with transgenic model N=61 comparisons, CI =confidence intervals.

Variable	SMD	95% CI	P value	Tau^2	I^2	R^2
APP	-0.04	-0.65 to 0.57	0.0009	1.15	97.14	31.44
APP/PS1	-0.73	-1.54 to 0.09				
APP/PS1/MAPT	-1.79	-2.67 to -0.91				

Table 6.14: Univariable meta-regression analysis of number of rears with transgenic mode category
N=61 comparisons, p=0.0009. SE = standard error of the SMD effect size estimate, CI =confidence intervals.

Variable	SMD	95% CI	P value	Tau^2	I^2	R^2
Intercept	-1.99	-3.31 to -0.67	0.004	1.57	98.28	12.08
Height of test arena	0.04	0 to 0.08				

Table 6.15: Univariable meta-regression analysis of number of rears with height of test arena
N=43 comparisons, CI =confidence intervals.

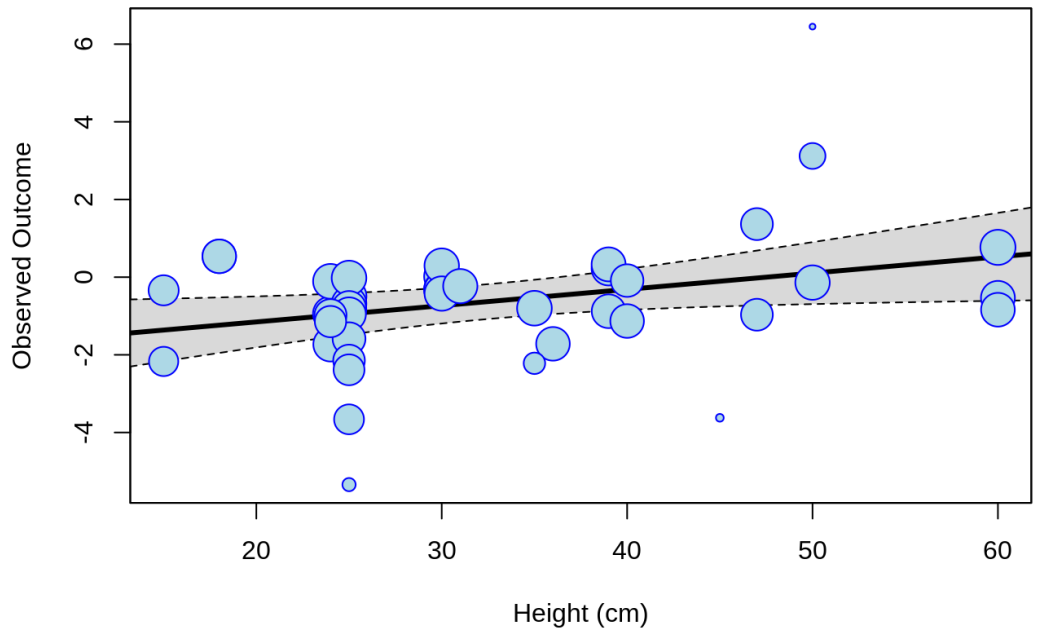


Figure 6.13: Bubble plot of meta-regression analysis with height of OFT and SMD effect size for number of rears
Black line indicates regression line. Grey shaded area represents confidence interval bounds. Bubble sizes reflect the weight of a modelling comparison. Positive SMD indicates a higher number of rears, N=43 comparisons.

4. ***Line/grid crossings***

In experiments measuring grid/line crossings in the open field. the pooled effect of modelling interventions was -0.13 (95% CI -0.40 to 0.14, n=58 comparisons). These results indicate that transgenic animals cross a fewer number of lines / grids in the open field than wild-type animals. Heterogeneity was high (I^2 = 97.4%, Q=2202.1, Tau^2=1.0). A forest plot of comparisons is shown in Figure 6.14.

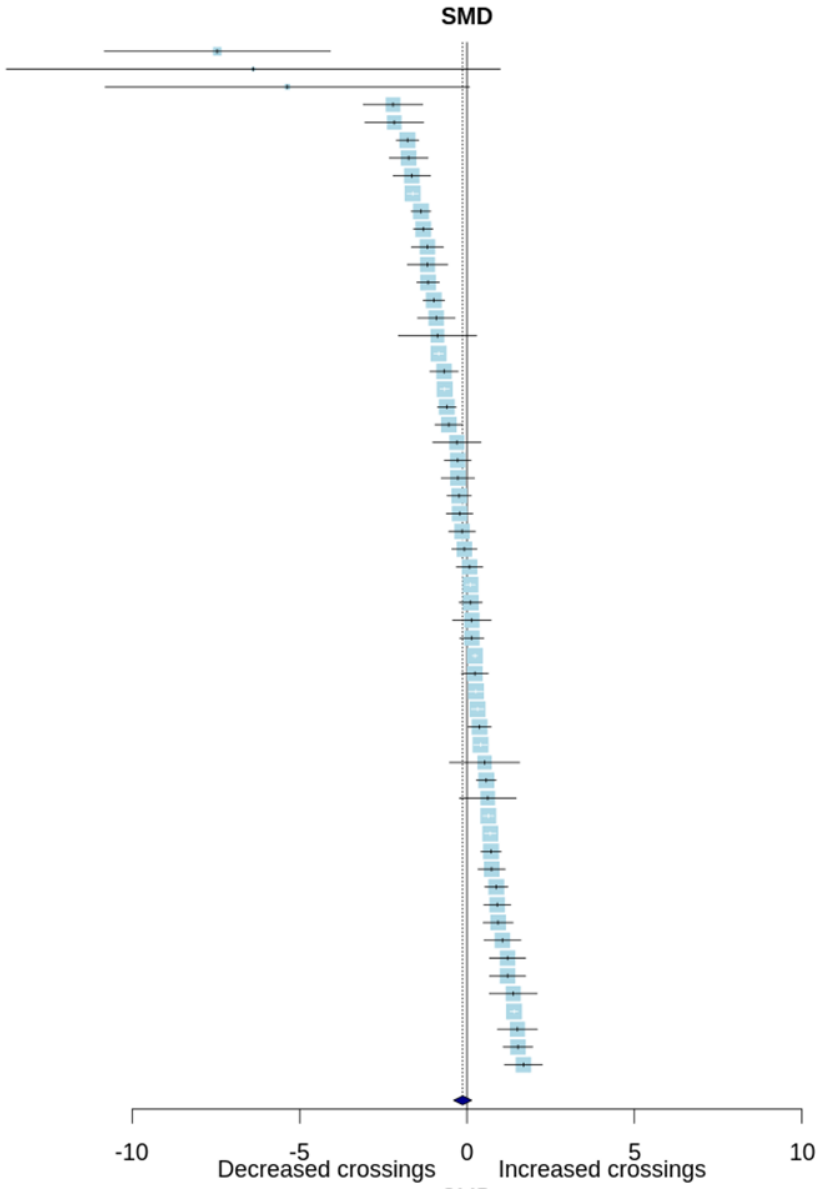


Figure 6.14: Forest plot of effect sizes for number grid/line crossings in the OFT
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. A high resolution size image with more detail is available at: https://osf.io/fc6kd/

Several variables were significant in the univariable analysis (Tables 6.16 – 6.20). Higher grid/line crossings (increased SMD estimates) were observed in Tg2576 models and in the wider APP model category. Experiments in 3xTg-AD models and the wider APP/PS1/MAPT category had significantly lower grid/line crossings

(reduced SMD estimates). In studies where blinding was not reported, this was associated with a lower SMD estimate (fewer line/grid crossings). Experiments with white walls had lower SMD estimates (fewer line/grid crossings), while experiments with other wall colours had increased line/grid crossings. Where strain was not reported, this was associated with a lower estimate of effect (fewer line/grid crossings).

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
APP	0.49	0.17 to 0.81	<0.001	0.49	96.01	51.63
APP/PS1	0.14	-0.39 to 0.67				
APP/PS1/MAPT	-1.23	-1.74 to -0.71				

Table 6.16: Univariable meta- regression analysis of line/grid crossings with transgenic model category
N=58 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
3xTg-AD	-1.23	-1.63 to -0.83	<0.001	0.49	96.04	51.14
Other	0.28	-0.23 to 0.78				
Tg2576	0.59	-0.05 to 1.23				

Table 6.17: Univariable meta-regression analysis of line/grid crossings with transgenic model
N=58 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Blinding: Not Reported	-0.32	-0.65 to 0.01	0.015	0.88	97.73	11.99
Blinding: Reported	0.54	-0.15 to 1.18				

Table 6.18: Univariable meta-regression analysis of line/grid crossings with blinded outcome assessment.
N=58 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Wall colour: Black	0.1	-0.42 to 0.63	0.0004	0.69	96.81	31.1
Wall colour: NR	-0.76	-1.32 to 0.04				
Wall colour: Other	0.93	0.15 to 1.71				
Wall colour: White	-0.48	-1.26 to 0.3				

Table 6.19: Univariable meta-regression of analysis of line/grid crossings with arena wall colour.
N=58 comparisons, CI =confidence intervals.

Variable	SMD Estimate	95% CI	P value	Tau^2	I^2	R^2
Strain: C57BL/6	0.3	-0.29 to 0.89	0.0007	0.75	97.27	24.96
Strain: NR	-0.74	-1.46 to -0.03				
Strain: Other	0.38	-0.35 to 1.12				

Table 6.20: Univariable meta-regression analysis of line/grid crossings with background strain
N=58 comparisons, CI =confidence intervals.

In the model building exercise, I started with transgenic model category as a predictor as it explained slightly more of the heterogeneity than transgenic model. The final multivariable meta-regression model (Table 6.21) included transgenic model category, wall colour, and transgenic model as predictors. Together, the final model explained a significant proportion (R^2 = 65.15%, p<0.001) of heterogeneity in the dataset.

Variable	Regression weight (β coefficient)	95% CI	P value	Tau^2	I^2	R^2
(Intercept)	0.38	-0.07 to 0.82	<0.0001	0.45	93.28	65.15
APP/PS1	-0.26	-0.82 to 0.30				
APP/PS1/MAPT	-2.3	-3.15 to -1.45				
Wall colour: NR	0.55	-0.15 to 1.26				
Wall colour: Other	1.18	0.54 to 1.82				
Wall colour: White	0.94	0.14 to 1.75				
Model: Other	-0.68	-1.35 to 0.01				

Table 6.21: Multivariable meta-regression model for line/grid crossings

N=58 comparisons, CI= confidence intervals

5. ***Distance in centre***

In experiments measuring distance in the centre area of the open field. the pooled effect of modelling interventions was -0.23 (95% CI -0.59 to 0.13, n=30 comparisons). Transgenic animals travelled a shorter distance in the centre area of the open field compared to wild-type animals. Heterogeneity was high (I2 = 96.4%, Q=795.76, Tau2=0.93). A forest plot of comparisons is shown in Figure 6.15.

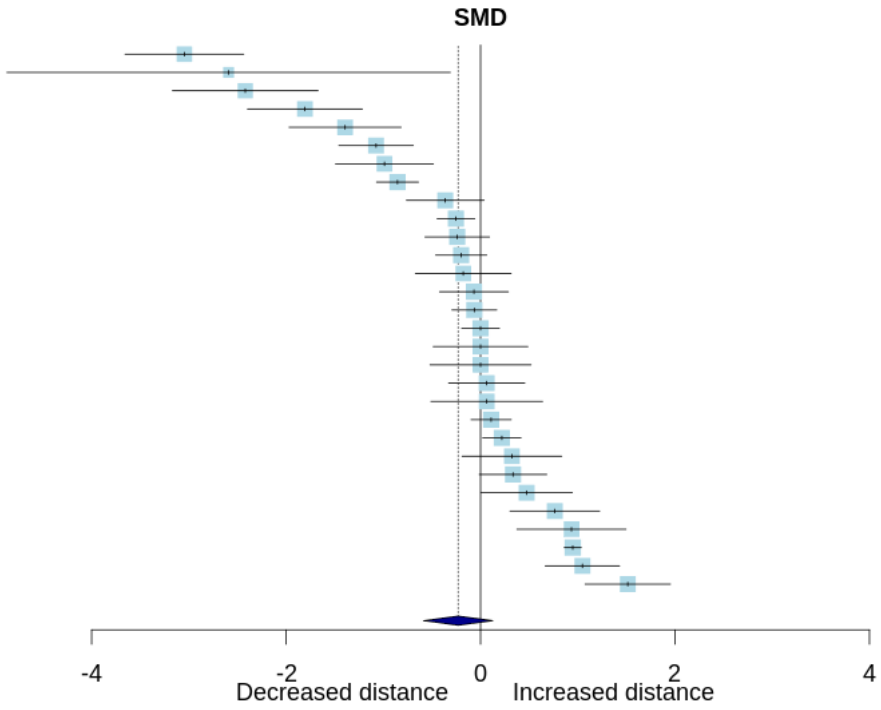


Figure 6.15: Forest plot of effect sizes for distance travelled in centre of open field. The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. A high resolution size image with more detail is available at: https://osf.io/fmd2c/

In the univariable analysis, blinded outcome assessment explained a significant proportion of the heterogeneity (Table 6.22). Blinded studies had a higher SMD estimate of effect (longer distance travelled in centre). Initially, in the model building exercise, the addition of many variables significantly reduced the AIC and it continued to decrease as more significant variables were added. I believed there

was a serious risk of overfitting due to the small sample size, so I re-built the model and checked each of the AIC reductions carefully. Following the addition of sex as a predictor, the next round of model fitting produced only minor changes in AIC (1-2 points) but these changes had a $p<0.05$. For this reason, I retained only blinded outcome assessment and sex as predictors in the model. Together, these variables explained 43.45% of the heterogeneity (Table 6.23).

Variable	Estimate (SE)	95% CI	P value	Tau2	I2	R2
Blinding: NR or No	-0.57	-1.04 to -0.11	0.0229	0.78	97	15.51
Blinding: Reported	0.27	-0.44 to 0.99				

Table 6.22: Univariable meta-regression analysis of distance in centre with blinded outcome assessment
N=30 comparisons, CI= confidence intervals

Variable	Regression weight (β coefficient)	95% CI	P value	Tau2	I2	R^2
Intercept	-0.63	-1.14 to -0.12	0.0063	0.52	95.56	43.45
Blinding: Reported	1.15	0.48 to 1.83				
Sex: Female	-0.45	-2.18 to 1.29				
Sex: Male	-0.59	-1.3 to 0.11				
Sex: NR	1.22	0.16 to 2.27				

Table 6.23: Multivariable meta-regression model for distance travelled in centre
N=30 comparisons, CI= confidence intervals

6. ***Time in periphery***

In experiments measuring time in the peripheral areas of the open field. the pooled effect of modelling interventions was 0.17 (95% CI -0.25 to 0.59, n=29 comparisons). This overall meta-analysis indicates that transgenic mice greater time in the periphery compared to wild-type mice. Heterogeneity was high (I2 = 93.8%, Q=450.63, Tau2=1.27). A forest plot of comparisons is shown in Figure 6.16.

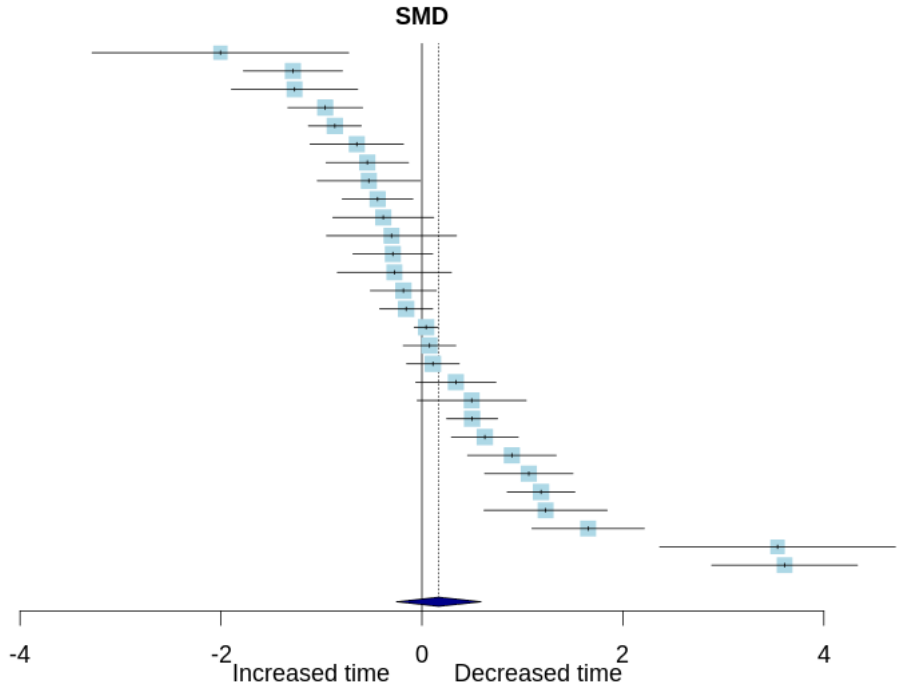


Figure 6.16: Forest plot of effect sizes for time in periphery.
The centre of each square represents the effect size of an individual comparison. Square size represents weight (%). Horizontal bars indicate 95% confidence intervals. The black diamond represents the pooled effect size estimate. A high resolution size image with more detail is available at: https://osf.io/4x5uk/

In the unvivartiate meta-regression (Table 6.24), wall colour explained a significant proportion of the heterogenetiy. However, it is not viable to draw any conclusions form this result. Wall colour was often not reported, there was not enough data to assess individual wall colours and data were grouped into one "other" category. No additional variables improved the model fit, so multivariable meta-regression was not possible.

Variable	SMD	95% CI	P value	Tau2	I2	R^2
Wall colour: NR	0.81	0.29 to 1.33	0.0009	0.81	95.88	36.05
Wall colour: Other	0.54	-1.28 to 0.2				

Table 6.24: Univariable meta-regression analysis of time in peripher with arena wall colour N=29 comparisons, CI= confidence intervals

6.5 Discussion

6.5.1 Overview of findings

This SR and meta-analysis of OFT outcomes measured in APP transgenic models summarised the literature across 237 publications describing modelling interventions and treatment interventions. I did not conduct a meta-analysis of treatment effects due no treatment being tested in more than four independent publications. Overall, the meta-analysis suggests that transgenic AD mice spend more time in the periphery, spend less time and move shorter distances in the centre, rear less, and cross fewer grids/lines in the OFT arena. Interestingly, in contrast with the hypoactive, anxious phenotype displayed across other outcomes, total distance travelled was increased in transgenic mice versus controls. However, the heterogeneity observed across experiments was very high.

6.5.2 Modelling of Alzheimer's disease in APP models

Across several outcomes, there is a pattern of hyperactivity in single APP transgenic mouse models, and hypoactivity in APP/PS1/MAPT mouse models. The activity estimates for APP/PS1 models are typically somewhere in the middle. There was often not enough data available to fully investigate specific mutations in greater detail, however complimentary results were obtained from analysis of individual transgenic models, and experiments in the Tg2576 model were associated with higher line/grid crossings and distance travelled. The CRDN8 model was also associated with higher estimates of distance travelled.

Wider 95% CIs were observed for larger effect sizes, where activity was markedly increased or decreased in transgengic animals versus controls. This is suggestive of small study effects, where experiments showing greater effects are of poorer methodological quality and/or are more likely to be published than experiments with smaller effect sizes. For line/grid crossings, it is interesting to note that this

variation is more evident at the decreasing end of the scale, which may be indicate a publication bias favouring studies which report decreased activity in transgenic models.

In contrast with other reports (Lalonde et al., 2012), I did not observe a clear age-dependent effect on hyper/hypo activity in the open field, even when accounting for transgenic model in the multivariable meta-regression. Although animal age at time of outcome measurement was largely reported, it did not explain a substantial proportion of the heterogeneity for any outcomes. Anecdotally, it should be noted that there were often large age ranges reported within papers reviewed. Age is typically reported in months and may range across two or three months. Due to the lifespan of a mouse, this age difference could be the difference between AD pathology developing and impacting upon phenotype or not. To adequately assess the OFT profile across different ages, age should be carefully controlled within each group to ensure animals are indeed at the same stage of AD pathology and reported in as much detail as possible to determine to exact age (i.e., in weeks or days, rather than months). Although treatment interventions were not the focus on this analysis, treatments were often administered prior to amyloid pathology.

Sex did not explain a significant proportion of the heterogeneity observed between studies for most outcomes, however, the lack of reporting and tendency to use only male animals makes this difficult to adequately assess.

6.5.3 Impact of OFT protocol on OFT outcome measures

For the most measured outcome, total distance travelled, combining variables in a multivariable meta-regression model did not explain any more of the heterogeneity. It is unclear which additional aspects of study design may be influencing results. I was not able to assess the impact of light intensity in any meta-regression model, as there were too many comparisons in respect of which this was not reported.

Illumination has been shown in multiple studies to impact upon ambulation (McReynolds, Weir, & DeFries, 1967; Nagy & Glaser, 1970; Trullas & Skolnick, 1993) and efforts should be made to raise awareness of this among researchers conducting such behavioural tests.

Furthermore, height of arena and area of the arena (calculated from a combination of length and width) were often missing in more than 10% of comparisons, excluding potentially relevant variables from being investigated further in the multivariable meta-regression. Wall colour was identified as a variable which explained a significant proportion of the heterogeneity across multiple outcome measures (time in centre area, line grid crossings, time in periphery). OFT arenas with white walls had lower estimates of time in centre and lower line/grid crossings. However, it is important to note that these observations may also be influenced by whether the animals were precluded from seeing the surrounding environment (opaque walls of any colour) or not (transparent walls). Nevertheless, given the potential impact of wall colour, it is concerning that 58.3% of publications did not report this parameter. It is possible that the material of the OFT arena also played a role here, but this was not a variable we extracted.

6.5.4 Internal validity

Risk of bias measures were often not reported. I was unable to assess the potential impact of power calculations as so few studies reported one (5/237). Less than 30% of relevant publications reported randomisation, and less than 25% reported blinded outcome assessment. A review conducted on the published preclinical AD literature in 2009 identified that no publications reported a power calculation, 16% of publications reported randomisation, 22% reported blinding, 13% reported a conflict-of-interest statement, and 56% reported compliance with animal welfare regulations (Egan et al., 2016). Based on these estimates, it is encouraging to note that there seems to have been some improvement in reporting quality over time.

There have been substantial research improvements since 2009, including the introduction of the ARRIVE guidelines (Kilkenny et al., 2010) However, most publications still fail to report the four key measures (randomisation, blinded outcome assessment, reporting of exclusions, and power calculations) considered most important for transparency in the reporting of animal research. This impedes quality assessment, as it remains unclear whether "not reported" means that experimenters did not make any attempt to reduce the risk of bias, whether there were reasons such measures could not be implemented, or whether experimenters did adopt measures to reduce the risk of bias but did not report it in the publication. There was a significant effect of blinding on the number of line/grid crossings and distance travelled in centre of the open field. Studies reporting blinded outcome assessment were associated with a higher SMD estimate of effect (increased grid/line crosses, increased time in centre).

6.5.5 Limitations

There are several limitations to this review. Included publications were, in the most part, annotated by a single trained reviewer. I thoroughly sense-checked the dataset, and corrected errors where feasible by re-extracting the data or using an alternative reviewer's annotations when possible. Despite these efforts, it is likely that some errors remain, potentially impacting upon the results.

Due to the poor reporting of some potentially relevant variables (including OFT protocols and measures relating to study quality and risk of bias), there were limits on our ability to understand the impact each variable had on heterogeneity.

A further limitation is that I did not analyse assessments of OFT behaviours over time within the same animals. A limited number of studies did report repeated measurements from the same cohort of animals, and it is our intention to analyse this as part of the final, larger review of OFT across all AD animal models.

Using automated tools comes with its own limitation, as they may not be accurate 100% of the time. In this case, I used a text-mining approach to identify relevant publications within a wider dataset, with the criterion of more than three mentions of the open field test in the full-text. It is not feasible to go through every publication in the database to check for missed OFT data, but it is my future intention to go through all publications which matched the regular expression 1-3 times.

6.5.6 Future directions

Having thoroughly reviewed the literature, I have identified additional variables which we did not extract, but which would be interesting to investigate in future reviews. As mentioned previously, the material of the OFT could also have an impact and may be highly related to the effect of wall colour that we observed. Another variable we did not assess is time acclimitised to the testing room, which may have an impact on rodent anxiety levels.

The OFT is often conducted as part of a battery of behavioural tests. It would be interesting to understand the influence of previous behavioural tests and identify whether there is an effect of habituation or reduced inhibition in studies where several tests are carried out on the same cohort of animals. Future work could also utilise the data from this review and incorporate findings from other activity-dependent behavioural tests. By integrating enough data from each distinct model, it may be possible to understand distinct Alzheimer's model phenotypes in greater detail. Activity levels will impact upon all behavioural tests, so it is fundamental to understand how we should expect a specific transgenic animal model to behave at a given point in time in the open field test and what the knock-on effect will be in other measures used to assess the impact of treatment interventions (e.g., the Morris water maze).

Recently, a human adaptation of the open field test was developed in an attempt to achieve cross-species translation (Gromer, Kiser, & Pauli, 2021). Humans appear to display similar thigmotaxis-like behaviour to rodents, although an attempt to map this behaviour onto anxiety traits was not successful. As the authors state, "Future studies should use the approach to further elaborate what exactly the open field test measures and under which conditions anxiety modulates human open field behaviour". Despite decades of research utilising the OFT in rodents, the same is true. Once the larger review is complete, I plan to use the data collected to perform a principal components analysis to identify the factors underlying behaviour in the OFT. Studies have already explored this on a much smaller scale using results from individual animal experiments (Tanaka et al., 2012) but using a meta-research approach I will have a much greater sample size, distributed across a range of studies.

6.6 Conclusions

This work examined the OFT literature in transgenic APP models of Alzheimer's disease. Key measures to reduce the risk of bias were not reported in most publications, casting doubt on the validity of findings. Furthermore, important methodological details about the open field arena were often not reported – including the size, shape, illumination, and wall colour. Most OFT outcomes were dependent on the transgenic model used. There was a general pattern observed across multiple activity-based outcomes that single APP transgenic mouse models displayed higher motor activity, APP/PS1 models displayed a moderate amount, and APP/PS1/MAPT moved least. There was substantial between-study heterogeneity for each outcome and for the most reported outcome (total distance travelled) this was largely unexplained. Using a multiple meta-regression approach, I was able to explain a large proportion of the variability in line/grid crossings (horizontal activity) in the open field, with model category, wall colour, and transgenic model explaining over 65% of the heterogeneity. This work can inform the robust experimental

design and reporting of future studies, within both the Alzheimer's literature and beyond. The open field test is a widely used but little understood behavioural measure. More large-scale efforts should be made to understand the biological meaning of individual OFT measures and the usefulness of the test in understanding disease model phenotypes and general motor activity across animal species, strain, age, and sex.

CHAPTER 7: APPLYING SOLES METHODOLOGIES TO FACILITATE EVIDENCE SYNTHESIS IN THE COVID-19 PANDEMIC

7.1 Chapter Introduction

In this chapter, I describe the application and development of the SOLES methodology described in Chapter 2 to primary COVID-19 research. This work began in March 2020 in response to the global pandemic. As a research group, we believed we had useful expertise to contribute and wanted to develop a tool to aid in synthesising the masses of research evidence emerging on the SARS-CoV-2 virus and COVID-19 disease.

Building COVID-SOLES was a collaborative team effort. My primary role within the project has been to develop a living, automated (where possible) pipeline to collect and visualise the COVID-19 literature, building on my experience with AD-SOLES. This project required that I acquire a set of new skills. I learned to build and write to Structured Query Language (SQL) databases, connect to APIs, understand complexities around character encoding, scheduling R scripts to run at specified times, and document complex programmatic workflows.

This work demonstrates the potential of the SOLES approach in an entirely new research area. I plan to use the data science skillset I have developed throughout this project to redesign and improve upon the pilot AD-SOLES and deliver a curated resource which synthesises the preclinical AD literature as and when it emerges.

7.2 Background

The COVID-19 pandemic has changed our personal and professional lives significantly. Since SARS-CoV-2 first emerged in China's Hubei Province, scientists across the globe have been working around the clock to understand as much as

possible about the virus (ie how it spreads, how it enters the body, and how it impacts health, society, and the economy).

This focus has brought with it a substantial influx of preprints and publications. Many are no doubt extremely useful and insightful and some, perhaps, capitalise on the intense spotlight on COVID-19 related commentaries and reviews. Even with dedicated “Coronavirus collections” on publisher webpages, the amount of research emerging each day is almost impossible to keep up with.

With these issues in mind, and a desire to contribute our expertise towards global efforts to characterise and tackle the virus, our research group set out to create a systematic online living evidence summary (SOLES, see Chapter 2) of COVID-19 research. It was our intention to create a resource which synthesised the evidence in a useful way for researchers working in the field, for systematic reviewers, and for decision-makers - including governments and funding bodies.

We pre-registered a protocol for COVID-SOLES on the Open Science Framework (Appendix C7.1). We developed a search strategy to identify COVID-19 publications across several databases and continue to retrieve new publications on a weekly basis. The COVID-SOLES workflow is largely R based and uses code which I have written. First, citations retrieved from databases are formatted and combined. For PubMed, citations are retrieved automatically using their API while other databases are searched manually. In addition, I created a MySQL database to manage citations as they come into the workflow, and drafted custom R code to interact with this. Using ASySD (Chapter 4), I removed duplicate publications and preferentially retained the older version of each record already present in the database for data consistency.

To annotate COVID-19 publications, we recruited and trained a team of reviewers and created a SyRF project with annotation questions, asking reviewers about the

study and, if it was a primary study, further questions about objective, methodology, and subjects/samples tested. Once a reviewer passed the training, they could annotate new publications. Each publication was reviewed by two independent reviewers and reconciled by a third in the event of disagreements. The final annotations were used to classify each paper in a useful way. When thousands of studies had been classified, we were able to start adopting ML methods into our workflow to pass only primary research onto reviewers for further classification.

Finally, I created an interactive R Shiny web application (see Appendix C7.2) for users to interact with the COVID-SOLES database, track annotation progress, and download annotated datasets of publications. This work was subsequently published in the Journal of the European Association for Health Information and Libraries.

7.3 Publication

Feature Article

Building a Systematic Online Living Evidence Summary of COVID-19 Research

The CAMARADES COVID-SOLES Group*
Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK
Kaitlyn.Hair@ed.ac.uk

*The members of the CAMARADES COVID-SOLES Group are reported before the References

Abstract

Throughout the global coronavirus pandemic, we have seen an unprecedented volume of COVID-19 research publications. This vast body of evidence continues to grow, making it difficult for research users to keep up with the pace of evolving research findings. To enable the synthesis of this evidence for timely use by researchers, policymakers, and other stakeholders, we developed an automated workflow to collect, categorise, and visualise the evidence from primary COVID-19 research studies. We trained a crowd of volunteer reviewers to annotate studies by relevance to COVID-19, study objectives, and methodological approaches. Using these human decisions, we are training machine learning classifiers and applying text-mining tools to continually categorise the findings and evaluate the quality of COVID-19 evidence.

Key words: *COVID-19; evidence synthesis; machine learning; web application; database.*

Background

The COVID-19 pandemic continues to present a major challenge for health services and society worldwide. Since the emergence of the SARS-CoV-2 virus, the research community has shown an extraordinary response to the pandemic. This volume of information and rate of publication makes it exceedingly challenging for research stakeholders (including researchers, funders, and policymakers) to efficiently identify studies relevant to their interests, evaluate the quality of those studies, and utilise their findings for health benefit (1). This "infodemic", along with the dissemination of unsubstantiated claims in both lay and social media, risks fuelling a growing distrust in science and highlights the need for an accessible resource to support public understanding of, and access to, research findings.

Evidence is incremental, and new experimental findings offer the greatest value when considered in the context of other studies that have addressed the same or related research questions in different settings. Systematic reviews capture, summarise, and critically appraise the available evidence relevant to a pre-specified research question. They are considered the most effective method of reaching a rigorous understanding of the literature, and informing decision-making (2). Unfortunately, the time taken to perform traditional systematic reviews means that the findings are often outdated by the time of dissemination. The urgent need for evidence-based treatments for COVID-19 infection combined with a rapidly accumulating COVID-19 literature has made this an even greater challenge. Automation technologies (e.g. machine learning and text-mining) can be used to reduce the time and resources required. For example, we can train a machine to classify research as relevant or not relevant to our research question, or to extract structured information from publications, at greatly reduced human effort (3-5). Such technologies facilitate the development of "Living" systematic reviews, in which new evidence is incorporated into the review as and when it becomes available (6, 7). Further, by incorporating crowdsourcing approaches to recruit and train external reviewers, a much larger team can work together to extract information from publications at a faster pace.

Building upon existing living review methodologies, we have developed and integrated a series of automation tools and methodologies for the continual collection, categorisation, and quality assessment of COVID-19 evidence from primary research studies. We have built a Systematic Online Living Evidence Summary (SOLES) of all primary research relevant to COVID-

Address for correspondence: Kaitlyn Hair, the CAMARADES COVID-SOLES Group, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh BioQuarter, 49 Little France Crescent, Edinburgh, EH16 4SB, UK.
E-mail: Kaitlyn.Hair@ed.ac.uk

19; an interactive web application, which allows users to interact with a visual summary of the curated information, interrogate the dataset, and download relevant citations filtered by study characteristic of interest. This resource is intended for use by all stakeholders in COVID-19 research, including researchers working within the field or performing rapid or systematic reviews of COVID-19 literature.

METHODS

Identifying new research papers

To retrieve up-to-date research reports we retrieve citations weekly from PubMed (National Library of Medicine), Web of Science (all available databases: Web of Science Core Collection, BIOSIS Citation Index, Current Contents Connect, Data Citation Index, Derwent Innovations Index, KCI-Korean Journal Database, MEDLINE, Russian Science Citation Index, SciELO Citation Index, Zoological Record), EMBASE (OVID), and the World Health Organisation's COVID-19 database (8). Our search terms are described in our study protocol and have been updated over time to address changes in COVID-19 research terminology (9). To identify new research from PubMed programmatically, we use the pubmedTools R package (10) developed within our group to access the Entrez application programming interface, while other records are obtained through manual searching of the platforms/databases outlined above .

Duplicate removal

To maintain a database of unique citations, we identify and remove duplicate citations (bibliographic duplicates of work published in the same journal at the same time by the same authors) identified across different databases using an automated, R-based tool developed within our research group, the automated systematic search de-duplicator (11).

Retrieving full text publications

We retrieve full-text publications using custom R code (12) to access full-text portable document formats (PDFs) where we have institutional access (University of Edinburgh). The extraction code uses digital object identifiers (DOIs) to retrieve PDF links through CrossRef, PubMed Central, and doi.org, then downloads the PDF file using the retrieved link.

Crowdsourced study annotation

To adequately capture the broad spectrum of primary COVID-19 research, we developed a schema (*Figure 1*) to classify research by type, objective, methods, and patient population/ sample type, based on previously proposed definitions (13). Using these classifications, we

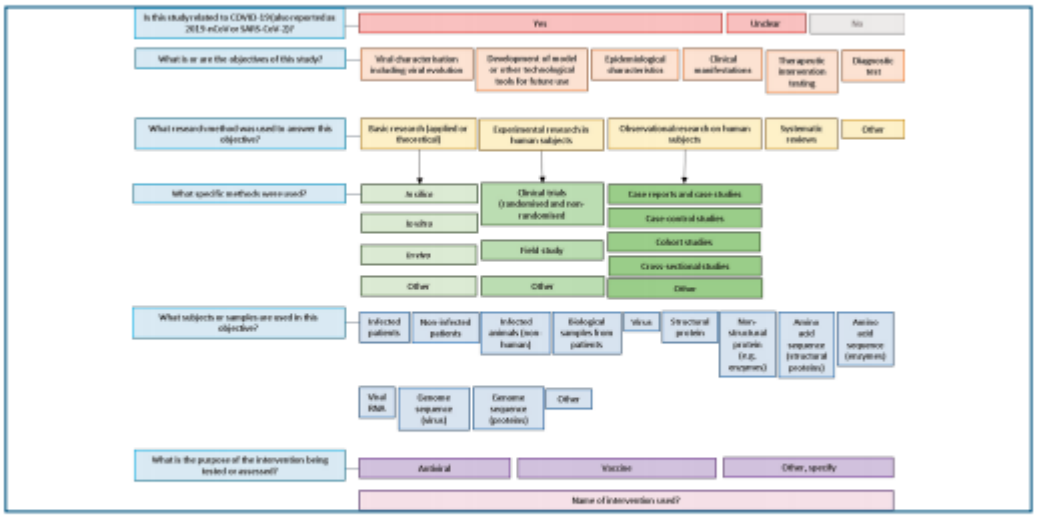


Fig. 1. *Research classification schema for primary COVID-19 studies. Arrows indicate a tree-like structure where reviews can only add subsequent annotations based on the previous annotation.*

designed a project on the Systematic Review Facility (SyRF; http://syrf.org.uk/), a widely used and freely available online platform developed within our research group (14). SyRF facilitates the conduct of large, collaborative systematic review projects and allows users to design structured annotation forms with custom questions. Once the project plan had been finalised, three independent researchers within our group annotated a test batch of 16 research papers. Through discussion, we arrived at a consensus on how each paper should be annotated. These annotations became our "gold-standard" annotated dataset used to train a crowdsourced team of human reviewers.

To recruit a team of reviewers to annotate COVID-19 research, we advertised the project via our social media profiles, existing contacts, and university research networks. Trainee reviewers were required to annotate a minimum of eight papers which were then checked against the gold-standard annotations. Once complete, we provided feedback and either asked trainees to complete more training papers or allowed them to continue as a reviewer on the main project. To ensure quality, each article is annotated by two independent reviewers. To keep reviewers up to date, fortnightly progress reports are sent out via email. Reports are generated programmatically with R code which interacts with SyRF and published online on the RPubs server as a living RMarkdown document (15).

Integration with the Systematic Review Facility

Subsets from our dataset of unique COVID-19 records are selected based on the date they are retrieved, with older records uploaded first. Custom R scripts are scheduled (using the CronR package) to periodically interact with SyRF to obtain information on the number of reviewers working on the project, the number of studies annotated, and the annotations themselves. This allows us to keep an up-to-date record of progress.

Reconciliation of annotations

For each paper, annotations from two independent reviewers are compared using a custom R script. If reviewers agree on whether the paper describes primary research relevant to COVID-19, this study is immediately classified as "included" or "excluded" – irrespective of whether they agree on all classifications. If reviewers do agree across all classifications, the study is classed as "reconciled" and those classifications are final. If there are disagreements on one or more annotations, the paper is passed to a senior reviewer who will reconcile the disagreements before submitting a final set of classifications.

Machine-assisted classification of primary studies

We used the "included" or "excluded" decisions from reconciled annotations to train a machine learning algorithm hosted by collaborators at The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), University College London. The algorithm uses natural language processing to identify features within the Title and Abstract of citations. We aimed to train it to automatically classify non-annotated studies as either "primary COVID-19 research" or "other" research.

Web application and dataset availability

We built a user interface to access our entire COVID-19 dataset via an R Shiny web application. The application allows users to visualise the annotated evidence, search the citation database (using regular expressions), and download relevant citations. The COVID-SOLES application is freely available online (16).

RESULTS

COVID-SOLES citation database

At the time of writing (May 2021) we have identified a total of 812,261 potentially relevant citations since our COVID-19 searches began in March 2020. The distribution of records retrieved from each database is shown in *Figure 2*. We obtained the highest number of records from the WHO COVID-19 database (N= 246,299) and the lowest number from PubMed (N=129,973).

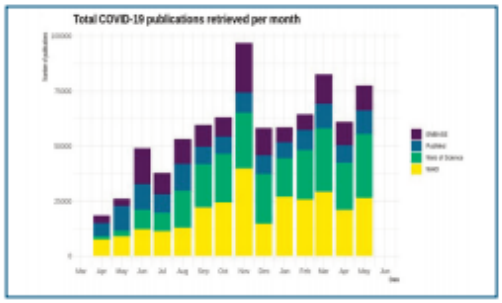


Fig. 2. *Total COVID-19 citations retrieved from each database per month.*

Following extensive de-duplication 349,726 unique citations have been identified. Over time, the number of unique publications retrieved per month has increased, with a brief levelling off period over the new year. In May 2021, we identified 50,095 publications, the largest monthly publication count yet.

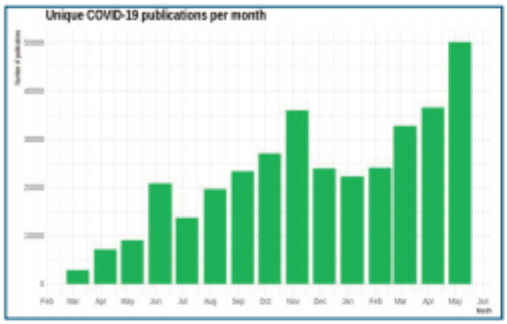


Fig. 3. *COVID-19 citations retrieved per month. Bars indicate unique citations retrieved across databases following the removal of duplicates.*

Crowdsourced annotation

We have recruited 88 trainee reviewers of which 78 have completed training and are able to annotate COVID-19 publications. The median number of papers annotated by each reviewer was 99 (interquartile range: 70.75 – 173.75). Two reviewers were particularly active, annotating 1,874 and 6,597 publications, respectively.

Machine classification of COVID-19 research

From a total of 226,417 citations in our dataset which had abstracts, 3,405 had been classified by humans as "primary COVID-19 research" (N=1312) or "other" (N=2093). This dataset was randomly split into a training set, validation set, and test set. We used a pre-set sensitivity threshold of at least >95% to ensure we captured the majority of relevant publications. On the test set (N=681), the classifier performed at a sensitivity (percentage of citations correctly included) of 95.2%, a specificity (percentage of citations correctly excluded) of 76.6%, and precision (percentage of correctly included citations from all included citations) of 71.9%. To date, the number of fully annotated primary studies is too low to train classifiers to identify specific objectives or study methodologies (N=1,174). A summary of the primary studies annotated by objective and methodology is shown in *Figure 4*. Due to our chronological approach to annotating studies, this summary reflects COVID-19 research conducted early in the pandemic, in March and April 2020.

Use of Web application

Since we developed the COVID-SOLES application, it has been accessed over 1,700 times by users from 45 countries.

LIMITATIONS AND FUTURE WORK

Optimising citation retrieval

Some retrieved citations lack useful meta-data, such as

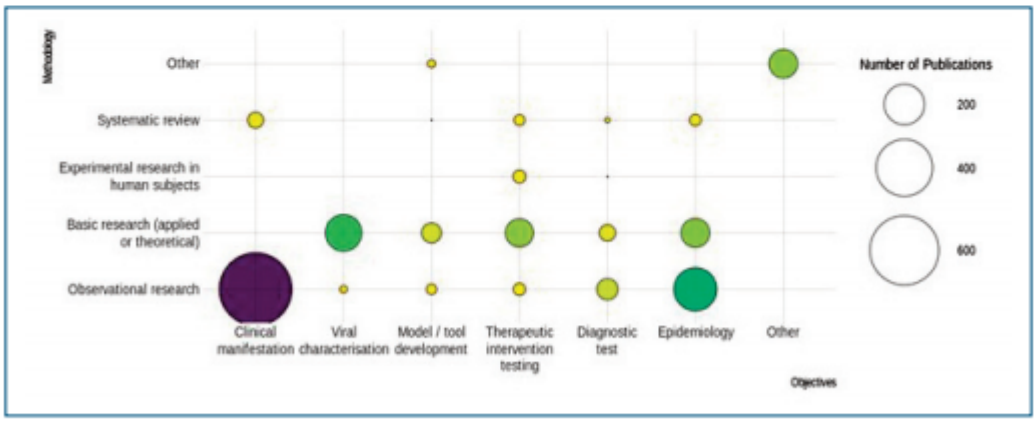


Fig. 4. *COVID-SOLES database citations annotated by objective and methodology (N=1,174). Darker colours and larger bubble sizes indicate a higher number of publications.*

DOI. This may be, in part, due to the uniquely challenging pace of COVID-19 research and our continual searching to retrieve newly published research. In some cases, we may be retrieving publications before they are fully indexed in biomedical databases. *Figure 5* indicates the percentage of unique citations retrieved from each database that lacks digital object identifiers (DOIs). Of unique records retrieved from the WHO COVID-19 database and Web of Science, 33.5% and 21.3% of citations are missing DOIs, respectively. To remedy this, we are now employing the rcrossref R package (17) to programmatically query the CrossRef database using titles and to identify the corresponding DOI information. Furthermore, we are refining our deduplication code to set a preference for retaining PubMed records over other databases, as 95.8% of citations we receive from PubMed have DOIs.

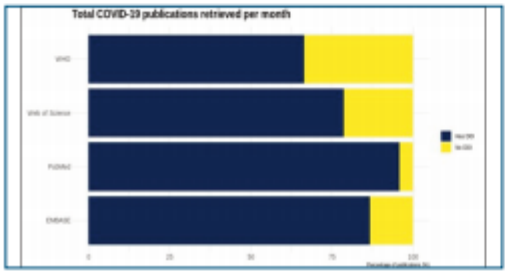


Fig. 5. *DOI status across databases searched to obtain COVID-19 citations.*

Supplementing our study type annotations

A major limitation is that we are not yet able to classify research automatically. The ability to do this as new research emerges would provide us with insights into research trends over time and identify gaps where more research may be needed. To obtain more study type annotations to drive automatic study type detection, we aim to recruit more volunteers by launching a new campaign across social media and other research networks. We are also exploring the possibility of exploiting annotation data from other openly available systematic evidence summaries of the COVID-19 literature and from published systematic reviews with accessible data. Past reviews have focused primarily on the clinical literature, so we will aim to make use of the existing data to classify human research and focus our crowd towards areas where there has been comparatively less attention e.g. in vivo research and in vitro research.

Improving our user interface

At present, some elements of the R Shiny user interface load slowly and it does not support full text searching of PDFs or Boolean searching of our database. We are currently building a new web interface to support these functionalities and sustain the growing COVID-SOLES database going forward.

Conclusion

We have developed a living workflow to synthesise COVID-19 research which enables research users to make rapid use of the currently available evidence. The SOLES workflow is sustainable, requiring minimal human effort to maintain – except the efforts of crowdsourced volunteers – and is transferrable to other research areas. We will continue to improve upon this workflow, enable more automated categorisation tools, and upgrade the user interface to enable features most useful to the evidence synthesis community.

Acknowledgements

This paper originates from a presentation at the International Collaboration for the Automation of Systematic Reviews (ICASR) meeting held in April 2021.

Members of The COVID-SOLES Group (authorship)

Project administration: Emily Sena (University of Edinburgh). **Study design:** Gillian Currie, Zsanett Bahor, Emily Sena, Malcolm Macleod, Emma Wilson, Kaitlyn Hair. **Software and architecture:** Kaitlyn Hair, Emma Wilson, Chris Sena, Can Ayder, Jing Liao, Ezgi Tanriver Ayder. **Annotation:** Joly Ghanawi, Anthony Tsang, Anne Collins, Malcolm Macleod, Alice Carstairs, Sarah Antar, Katie Drax, Kleber Neves, Thomas Ottavi, Yoke Yue Chow, David Henry, Cigdem Selli, Mariam Fofana, Martina Rudnicki, Brendan Gabriel, Esther J. Pearl, Simran S. Kapoor, Julija Baginskaite, Santosh Shevade, Alexandria Chung, Marianna Antonia Przybylska, David E. Henshall, Karina Lôbo Hajdu, Sarah McCann, Catherine Sutherland, Tiago Lubiana Alves, Rachel Blacow, Rebecca J. Hood, Nadia Soliman, Alison Harris, Stephanie L. Swift, Torsten Rackoll, Nathalie Percie du Sert, Fergal Waldron, Magnus Macleod, Ruth Moulson, Juin W. Low, Kristiina Rannikmae, Kirsten Miller, Alexandra Bannach-Brown, Fiona Kerr, Harry L. Hebert, Sarah Gregory, Isaac William Shaw,

Alexander Christides, Mohammed Alawady, Robert Hillary, Alex Clark, Natasha Jayasuriya, Samantha Sives, Ahmed Nazzal, Nimesh Jayasuriya, Michael Sewell, Rita Bertani, Helen Fielding, Broc Drury. **Senior reviewers (Reconciliation):** Joly Ghanawi, Anthony Tsang, Gillian Currie, Zsanett Bahor, Malcolm Macleod, Emily Sena.

Submitted on invitation.
Accepted on 10 June 2021.

REFERENCES

1. Palayew A, Norgaard O, Safreed-Harmon K, Andersen TH, Rasmussen LN, Lazarus JV. Pandemic publishing poses a new COVID-19 challenge. Nature Human Behaviour. 2020;4(7):666-9. doi: 10.1038/s41562-020-0911-0.
2. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, Al-Shahi Salman R, Chan AW, Glasziou P. Biomedical research: increasing value, reducing waste. Lancet. 2014;383(9912):101-4. Epub 2014/01/15. doi: 10.1016/s0140-6736(13)62329-6.
3. Bannach-Brown A, Przybyła P, Thomas J, Rice ASC, Ananiadou S, Liao J, Macleod MR. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Systematic Reviews. 2019;8(1):23. doi: 10.1186/s13643-019-0942-7.
4. Liao J, Ananiadou S, Currie GL, Howard BE, Rice A, Sena ES, Thomas J, Varghese A, Macleod MR. Automation of citation screening in pre-clinical systematic reviews. bioRxiv. 2018. doi: 10.1101/280131.
5. Bahor Z, Liao J, Macleod MR, Bannach-Brown A, McCann SK, Wever KE, Thomas J, Ottavi T, Howells DW, Rice A, Ananiadou S, Sena E. Risk of bias reporting in the recent animal focal cerebral ischaemia literature. Clinical science (London, England : 1979). 2017;131(20):2525-32. Epub 2017/10/14. doi: 10.1042/cs20160722. PubMed PMID: 29026002.
6. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, Salanti G, Meerpohl J, MacLehose H, Hilton J, Tovey D, Shemilt I, Thomas J. Living systematic review: 1. Introduction—the why, what, when, and how. Journal of Clinical Epidemiology. 2017;91:23-30. doi: https://doi.org/10.1016/j.jclinepi.2017.08.010.
7. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, Glasziou P, Shemilt I, Synnot A, Turner T, Elliott J. Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology. 2017;91:31-7. doi: https://doi.org/10.1016/j.jclinepi.2017.08.011.
8. WHO. Global literature on coronavirus disease. Available from: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/.
9. Currie G, Bahor Z, Liao J, Sena C, Hair K, Wilson E, Wang Q, Bannach-Brown A, Tanriver-Ayder E, Ayder C. Protocol for a "living" evidence summary of primary research related to Covid-19.
10. Liao J. pubmedTools 2021.
11. Hair K, Bahor Z, Liao J, Macleod MR, Sena ES. The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. bioRxiv. 2021:2021.05.04.442412. doi: 10.1101/2021.05.04.442412.
12. Liao J. fulltext_extractor 2020. Available from: https://github.com/shihikoo/fulltext_extractor.
13. Röhrig B, du Prel J-B, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009;106(15):262-8. doi: 10.3238/arztebl.2009.0262. PubMed PMID: 19547627.
14. Bahor Z, Liao J, Currie G, Ayder C, Macleod M, McCann SK, Bannach-Brown A, Wever K, Soliman N, Wang Q, Doran-Constant L, Young L, Sena ES, Sena C. Development and uptake of an online systematic review platform: the early years of the CAMARADES Systematic Review Facility (SyRF). BMJ Open Science. 2021;5(1):e100103. doi: 10.1136/bmjos-2020-100103.
15. CAMARADES. COVID-SOLES Progress Report. Available from: https://rpubs.com/CAMARADES/covid-soles-progress-report.
16. CAMARADES. COVID-SOLES. Available from: https://camarades.shinyapps.io/COVID-19-SOLES/.
17. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. rcrossref. Available from: https://github.com/ropensci/rcrossref.

This paper is published under a CC BY license

7.4 Conclusion

In this work, we developed a semi-automated workflow to synthesise COVID-19 evidence from publications spanning *in vivo, in vitro*, and clinical research. The emergence of COVID-19 had a dramatic impact on the research landscape, as scientists across the globe turned their attention to understanding the SARS-CoV-2 virus. The rapidly expanding collection of research publications presented a unique challenge for research users (including healthcare professionals, researchers, and policy makers) to make sense of the currently available evidence. The need to make informed decisions quickly in a rapidly changing situation underlines the value of automated meta-research approaches, such as the SOLES approach.

COVID-19 is an emerging discipline, which means we were not immediately able to develop and apply regex dictionaries to classify research automatically by COVID-19 model or treatment intervention. However, in collaboration with subject experts, this likely could be achieved. Work to train ML algorithms to classify the literature by research type is ongoing but will require substantial amounts of training data. Crowdsourcing the annotation of publications was a useful approach but requires time investment and training. As each publication was annotated for study objectives, methods, and subjects/samples in duplicate, there were often small differences between reviewers which required reconciliation. Overall, it took longer to fully annotate a primary study than was expected. In future crowdsourced projects, breaking down annotations into smaller decisions (e.g. “animal study” or “in vitro study”) and having a fewer number of annotation questions (similar to work described in Chapter 2), may be more feasible and provide useful training data at a faster pace.

This work demonstrates the feasibility of the application of SOLES to other research domains. COVID-SOLES was developed within months, and our curated and continually updated dataset has been available online since May 2020. Automation

approaches developed and built upon throughout this thesis were instrumental to this approach. Furthermore, the data science skills developed throughout this project will facilitate work to update AD SOLES (Chapter 2) to a truly “living” evidence summary in the near future.

CHAPTER 8: GENERAL DISCUSSION AND CONCLUDING REMAKRS

8.1 Chapter introduction

In this thesis, I set out to create, build upon, and validate automation tools to aid in evidence synthesis for preclinical AD research. Further, using these tools, I have conducted two preclinical SR projects to gain a clearer understanding of the existing evidence and inform future translational research.

In this final chapter, based on my findings I will discuss the feasibility of implementing automation approaches for preclinical SRs, observations on the transparency and reporting quality of studies included in the SRs described, and concluding remarks on how the translation of AD research may be improved.

8.2 Keeping pace with preclinical AD research

In Chapter 2, I described my work to collate all the available experimental evidence using transgenic AD models. I identified over 26,000 publications which are likely to contain *in vivo* research in transgenic AD models. Based on this dataset, approximately 1,800 new papers containing relevant transgenic AD animal data are published each year. Given the rapid pace of evidence generation, it is likely not feasible for researchers to dedicate the time required to identify, digest, and make use of this information. AD researchers have collected a staggering amount of data, but there is currently no fit-for-purpose system to curate new evidence as it emerges. SRs are useful tools to evaluate the quantity and quality of existing evidence, but to identify and collate evidence from hundreds (or thousands) of publications requires substantial input from dedicated teams of reviewers. Using traditional SR approaches can mean that findings are often not disseminated quickly enough, and that SRs are not updated on a regular basis due to the effort involved.

8.3 Feasibility of automated approaches to data curation

Early in this thesis, I introduce approaches to automate steps of the SR process. In Chapter 2, I also introduce the concept of SOLES; a new approach, enabled by automation tools, to curate research evidence and accelerate the conduct of SRs. I piloted this approach in the preclinical AD literature, focussing on experiments in transgenic models. Although this approach is not yet optimal, I believe it has the potential to have a meaningful impact on the research cycle and can make it easier for researchers to get a sense of what is already done, how much confidence we should have in those findings, and what needs to be done next. The utility and drawbacks of the automated meta-research approaches I have used are discussed under the subheadings below.

8.3.1 Collecting all relevant research

Developing systematic searches for SRs is always a balancing act between trying to ensure that no relevant evidence is missed without capturing too many irrelevant publications to sift through. Using automation approaches allows for the widening of systematic searches to prioritise sensitivity (recall of relevant publications) over specificity (removal of irrelevant publications).

Theoretically, if a machine-assisted approach has enough high-quality training data and enough computational bandwidth, there is no limit on how many publications can be assessed for relevance. In Chapter 2, I performed a non-specific search to identify all AD research, then trained a ML algorithm to identify research conducted in transgenic AD models. The algorithm automatically screened over 150,000 publications for relevance; a volume which would likely not be feasible to complete by a human reviewer.

For the SR projects I describe later in this thesis (Chapters 5 and 6), I used regex search patterns to identify outcomes of relevance within the full-text of

publications. As shown in Chapter 3, outcomes measured in preclinical experiments are often not clearly described in the title or abstract of publications. Combined with my wider initial search, this approach likely led to the inclusion of many publications which may easily be missed using traditional SR approaches i.e. using highly specific systematic search strategies.

Finally, a highly sensitive search to retrieve as many relevant citations as possible comes with the caveat that many citations will have been identified more than once across different databases. With extremely large searches, this issue is amplified and difficult to manage using conventional deduplication tools. As described in Chapter 4, to tackle this issue I developed a novel, automated tool specifically designed to deal with large preclinical SR search datasets. This tool enabled me to identify tens of thousands of additional duplicate publications missed by other tools, without the need to look through the search dataset manually to find matching citations.

There are several limitations within the automated approaches I have outlined to collect all available evidence. For some publications, there was no abstract available which meant they could not be assessed by the ML classifier, while in other cases there was no full-text PDF available for further automated categorisation using regexes. These roadblocks could be cleared with open-access publishing and improvements to the way citations are indexed in biomedical databases (e.g. to include an abstracts for every publication).

The ML classifier was overly sensitive, meaning that some irrelevant publications will have been included. Over time, the performance could likely be improved with greater quantities of training data, annotated by human reviewers.

More generally, publication bias hinders the availability of useful data and I identified evidence of this within the preclinical AD literature (see Chapter 5).

Potentially useful data may not be sitting in a file drawer but may instead be online in an unknown location. Throughout this work, I focussed solely on the published, peer reviewed literature. However, it would be beneficial to integrate evidence from grey literature into future SOLES iterations. Furthermore, given the time taken to publish experimental research findings, preprints may also be a consideration for any attempt to collect all the relevant evidence.

8.3.2 Retrieving new research as it emerges

The ultimate goal of SOLES projects is to synthesise new research evidence as soon as it becomes available, thus preventing any delay in making use of that evidence to inform future research and guide decision making. In the AD-SOLES pilot project (Chapter 2), I have not yet implemented this feature. However, the COVID-SOLES project (Chapter 7) demonstrates the feasibility of semi-automated "living" search which retrieves new publications on a weekly basis. At present, I have only implemented a fully automated search for publications within PubMed (via their accessible API). Other APIs are available for other literature databases but attempts to automate other searches have been hampered by issues in retrieving citation data of the same quality (with abstracts and other meta-data present). At present, this step still requires some human input. For a future AD-SOLES iteration, it may be acceptable to only retrieve citations from PubMed and preprint servers such as Biorxiv (which also has an easily accessible API) on a daily basis, while running other database searches manually on a less frequent basis.

The retrieval of new citations on a regular basis also facilitates attempts to update SRs with minimal additional effort. The new deduplication tool I describe in Chapter 4 was also developed with this feature in mind. In the COVID-SOLES workflow (Chapter 7) there is a substantial amount of duplication across and within databases we collect citations from. Due to weekly searching, we also retrieve the same records multiple times. Using ASySD, we can specify that we want to automatically

remove newer versions of a citation, while preferentially retaining the original version of the citation in the database. This workflow ensures that citations are not overwritten, which would cause considerable data-linkage issues if, for example, the citation was already annotated in SyRF by reviewers.

Together, these approaches bring us a few steps closer to making use of new evidence at a much quicker pace.

8.3.3 Categorising research by MIO

SRs questions generally focus on specific MIO elements within publications (see Chapter 3). Finding relevant research which contains these features is not straightforward and as discussed, relevant details are often missing from the title and abstract. As part of the SOLES approach (Chapter 2), I aimed to assign MIO categorisations to publications within the preclinical AD literature using regex search dictionaries for animal models, interventions, and outcome measures. Overall, these approaches are time-saving and can be highly sensitive, especially if enough effort is placed into designing regex patterns to account for every common synonym and every deviation in punctuation. However, the way preclinical research publications are currently structured does not allow for accurate detection of MIO elements within every experiment. A publication may describe several experiments in different animal models, testing different treatments, and measuring those animals on different outcomes. Further, it is difficult to determine how many regex matches indicate that a MIO element is part of an experiment described within a publication, rather than a match to a reference to other research. These caveats suggest that while regex approaches are useful for MIO categorisation, they cannot be relied upon; a human reviewer is still required to read the paper thoroughly to understand the connection between MIO elements.

8.3.4 SOLES approaches in other domains

In Chapter 7, I applied my knowledge from creating a pilot AD-SOLES to an entirely different research domain. As COVID-19 is an emerging discipline, I did not have the benefit of existing resources to build text-mining dictionaries. However, I was able to build a structured SQL database to securely house the dataset and to build semi-automated pipelines in R code which fed into a continuously updated web application. This demonstrated the adaptability of the SOLES approach. In future, I hope to develop the concept further and engage with the AD research community and stakeholders on a wider level. Eventually, I aim to build a full AD-SOLES platform which meets the needs of research users, funders, meta-researchers, clinical trialists, patients and their advocates.

The key barriers which remain are the automated retrieval of relevant citations and full-texts from certain databases (e.g. Web of Science), the sheer volume of annotated data required to validate machine classifiers or text-mining tools, and the scalability of the final web application.

8.4 Applications: making sense of the evidence from transgenic AD models

Using automated tools, I obtained two datasets of evidence as the starting point for SRs and meta-analyses focussing on commonly used transgenic AD models tested on commonly used outcome measures: synaptic plasticity and transmission measured via *in vitro* field electrophysiology and motor activity in the OFT. Performing SRs of the literature has exposed several key areas for improvement and for consideration for laboratory researchers working in this domain. Overall, findings from the meta-analyses indicate that transgenic AD models had reductions

in LTP and I/O relationships. In the OFT meta-analysis, AD mice spent less time in the centre of the field, had fewer rears, travelled less in the centre of the field, spent more time in the periphery, and had fewer line crosses versus controls. In contrast to other outcome measures, transgenic AD mice had a higher estimate of total distance travelled.

8.4.1 Internal validity

Across both SRs there was moderate reporting of measures to reduce the risk of bias. Reporting of blinded outcome assessment was found to explain a significant proportion of the heterogeneity for some outcomes in both reviews. For example, studies which reported blinding had higher estimates of LTP for transgenic animals i.e. there was less evidence of synaptic plasticity deficits. Very few experiments reported the use of a sample size calculation. Therefore, there was not sufficient power for me investigate whether this had an impact on results. Similarly, reporting of exclusion criteria was poor and within the *in vitro* electrophysiology review, less than 10% of publications reported this information.

A key finding within the *in vitro* electrophysiology review was that over half of publications did not clearly report the number of animals the hippocampal slices were taken from in at least one of the experiments described. Experiments using hippocampal slices to perform electrophysiological assessments which compare animals that receive an *in vivo* intervention (the induction of a model or an *in vivo* treatment regimen) should report the number of animals from which slices are derived and the number of slices used. If a sample size is reported without any descriptive context, it is not possible to determine whether the value refers to slices or animals. There is also no accurate way to determine from the number of slices how many animals were used, as this varies between laboratories and will depend on the health of the slices and the age of the animals. To mitigate the risk of pseudoreplication in the meta-analysis, I could not use data from experiments

which did not clearly define the number of animals used. This issue severely limited the amount of data I could use, and the power I had to investigate sources of heterogeneity within the data. This finding has also exposed the potential that there is substantial pseudoreplication within the analysis of electrophysiology data, if the slice number has also been considered as the sample size within the analysis (Lazic et al., 2018). There is clearly a need to address this issue across the electrophysiology community and develop clearer guidance on experimental units.

8.4.2 External validity

Across both SRs, there were very few publications which assessed the impact of the same treatment intervention. Without measuring outcomes across different laboratories and different animal models, it is hard to predict if results would be generalisable. Co-ordinated efforts to track the quantity and quality of independent experiments testing a given treatment using AD-SOLES could pinpoint these gaps and promote improvements in measuring the external validity of findings.

Male mice were used more frequently than female mice across the assessed AD literature and very few publications reported data from male and female mice separately. Further, in the *in vitro* electrophysiology SR, I identified that a quarter of publications did not report the sex of the animals for at least one outcome. Evidence suggests that a large percentage of traits are sexually dimorphic in animal models (Karp et al., 2017). There is a growing recognition of sex bias in neuroscience research and calls to increase the number of studies in female mice and, ideally, for researchers to present data for male and female animals independently (Beery & Zucker, 2011). AD affects both male and female patients, with some evidence that females are disproportionally affected by condition (Ferretti et al., 2018). Sex did not explain a significant proportion of the heterogeneity for most outcomes described within the SRs, however, this was difficult to adequately assess due to reduced numbers of female mice. Future work

should, where possible, report data from both sexes to identify clear phenotypic differences in animal models and to assess whether interventions have the same effect in both sexes. When comparing reporting differences between the OFT and *in vitro* electrophysiology reviews, it could be that sex of animals is more frequently reported in behavioural experiments versus other outcomes.

For both SRs, the median age of animals at time of outcome measurement was between 6-8 months. In many cases, this is prior to the development of Aß neuropathology. The age of animals was not found to explain a significant proportion of the heterogeneity in any meta-analysis, but any age-dependent effects may also have been obscured by a lack of experiments in aged animals. There has been considerable variation reported in the lifespans of transgenic AD mice, which are influenced by sex and genotype (Rae & Brown, 2015). Transgenic AD mice typically have higher mortality rates and reduced lifespans, but there is a need to characterise age related impairments across models and sexes to improve the translational value of preclinical experiments. With a greater understanding of lifespan and AD related pathology over time, findings from preclinical experiments and clinical trials may be more comparable, as both species can be treated at similar stages of AD progression. At present, many experiments measure mice at early stages of AD, while clinical trials target patients displaying symptoms, who will already have established AD related brain changes.

Due to the relationship between synaptic loss and cognitive decline in human AD patients, it would be beneficial to understand if these pathologies are linked within AD animal models, as a measure of construct validity. However, this was difficult to ascertain from publications as few reported both cognitive behavioural measures (particularly MWM outcomes) and *in vitro* electrophysiology outcomes. Furthermore, where MWM outcomes were reported, it was often unclear whether the same cohort of animals had been sacrificed and used to measure synaptic plasticity. Future work should clearly state the order of outcomes measured

through use of a diagram or other means. The order of outcome measurement may also impact upon the results of outcomes measured at a later timepoint i.e. learning during the MWM may impact upon synaptic plasticity. Overall, to attain a clearer understanding of transgenic model phenotypes and how different aspects of AD are recapitulated, experiments should ideally focus on measuring clinically relevant outcomes in parallel to maximise our understanding of AD progression within specific models.

8.4.3 Electrophysiological parameters which impact results

This SR identified failures to report key electrophysiological parameters, which may have important impacts on the direction of effect. For example, the magnesium concentration of the recording solution was found to explain a significant proportion of the heterogeneity but had been reported in just over half of LTP experiments. Further, higher percentages of maximal I/O for pre-LTP baseline recordings and a higher number of stimulations to induce LTP each led to a larger modelling effect (greater reduction of LTP). Although most publications reported details of the type of stimulation used to induce LTP and the number of stimulations used, the stimulation strength used for baseline recordings and LTP induction (based on I/O percentage) were often omitted. Further, many studies did not report whether anaesthesia was used for animal sacrifice. However, studies which did report anaesthesia had smaller reductions in LTP versus those which did not report the use of anaesthesia.

Greater data sharing could greatly improve our ability to understand the variation observed across *in vitro* electrophysiology experiments. Extracting electrophysiology data from figures is time consuming and may be inaccurate due to the number of overlapping datapoints.

Interestingly, in a post-hoc analysis I identified that evidence for LTP deficits had increased over time. The meaning behind this is unclear and could represent time-sensitive improvements in the methodology used to detect changes in LTP. Efforts to improve the transparency of reporting electrophysiology methods could enable meta-researchers to pinpoint sources of between-study variability and guide further improvements in electrophysiology protocols.

8.4.4 OFT parameters which impact results

Across several OFT outcomes, wall colour of the open field arena had an influence on the results. For arenas with white walls, mice moved less and spent less time in the centre of the field. Given the importance of this parameter, it is concerning that less than half of publications reported the wall colour of the open field apparatus used. Combined with the considerable variation seen in arena shapes and dimensions, improved reporting of wall colour could help us understand the impact of different arena aesthetics on behaviour.

Due to incomplete reporting, there was not adequate power to evaluate the influence of several parameters, including light intensity and habituation time. Given the very high between study heterogeneity of OFT results, improving the transparency of reporting could enable researchers to better understand the sources of between study variation. The OFT is a commonly used behavioural measure of motor activity which can be used to inform findings across other behavioural tests.

8.5 Concluding remarks

In this thesis, I aimed to explore the use of automated meta-research approaches in preclinical AD research to benefit preclinical SRs. It is clear that automated

approaches can not only accelerate SRs, but also have the potential to also ensure they capture more of the available evidence. The SOLES approach I have described makes use of automation tools to synthesise and display vast amounts of preclinical experiments, making it easier to make use of, build upon, and evaluate existing research findings. By performing two SRs of the preclinical AD literature, I have identified failures to report measures to reduce the risk of bias and other key methodological details. If we cannot assess the extent to which experiments are at risk of systematic biases, we cannot determine how much trust we should have in their findings. If methodological details are not reported, we cannot easily replicate results and meta-researchers cannot perform secondary analyses to understand sources of variation. To conclude, these findings suggest that we need co-ordinated efforts to improve reporting standards for preclinical AD research to improve its utility, reproducibility, and translational value. By dedicating resources to the continual synthesis and monitoring of research data by implementing automation technologies, we can ensure we are best placed to maximise the potential of laboratory based research findings and lay the foundations for evidence-based translational success.

APPENDICES

Appendices are organised based on chapter numbers and therefore start at Chapter 2 (C2). Due to the volume of additional materials, many are provided as shortened hyperlinks rather than attached to the thesis document.

Appendix C2

1. ***Protocol for AD-SOLES approach:***
https://doi.org/10.17605/OSF.IO/QSVDW

2. ***Screening and annotation guide for reviewers:***
https://doi.org/10.17605/OSF.IO/Z9KD3

3. ***Reference list of the 26,627 publications included in preclinical AD dataset:***
https://osf.io/9s8kh/

4. ***Github repository containing R code for Shiny web application***
https://github.com/kaitlynhair/AD-SOLES

5. ***Regular expression dictionaries:*** *https://doi.org/10.17605/OSF.IO/TSAKH*

6. ***Shiny web application:*** *https://camarades.shinyapps.io/LivingEvidence_AD/*

Appendix C3

1. *Updated RoB regexes*

RoB	Regex
Randomisation	((?<!not)(\brandom(ly)?.{0,10}(assign|divid|treat|split|determin|receiv|alloc|subdiv|categor|select|plac|design|(re)?distrib|separat|sort|take|chose))|((?<!not)(assign|divid|treat|split|determin|receiv|[Aa]lloc|subdiv|categor|select|place|design|(re)?distrib|separat|sort|take|chose).(at)?.{0,10}?random))|(?<!not)(\brandomi[sz]((ed)|(ation))).(in|to)|(were|was).(performed|conducted)?.*?randomi[sz]ed manner|randomi[sz](ed|ation).{0,5}(of).{0,20}(animals|groups|mice|rats|fish)
Blinding	((?<!not)((\b(blind(ed)?|masked|naive).{1,3}(as)?to)|blind(ed|ly)|(blind|masked|naive).{1,3}(manner|eval|observ|investigat|rate|rati|experiment|research|test|quantif|cod|with respect to|method|analys|condition|score|operator|examiner|rate))|((perform(ed)?|count(ed)?|conduct(ed)?|genotype|cod(ed)?|test|carried out (in)?).{0,10}(blind))|\'blind\'|\"blind\")|((was|were|observer(s)?|experimenter(s)?|researcher(s)?|tester(s)?|rater(s)?|person(s)?|investigator(s)?|operator(s)?|examiner(s)?|kept).((unaware|not aware|without awareness|unrevealed|not revealed|blind|hidden|naive|masked).(of|to)?(.the)?.(experimenter(s)?|researcher(s)?|tester(s)?|rater(s)?|person(s)?|investigator(s)?|operator(s)?|examiner(s)?|identity|treatment|group|experimental|drug|intervention|genotyp|treatment)))
SSC	((minimum|planned|target|calculated a) sample size(of)?)|(((sample|group) size(s)?|number of animals (in the group|per group|in each group)?).(were|was|of|for|would)?.{0,25}(at least|calculat|determin|estimat|calculat|consider|suffic|based on|for group assignment were made a priori|and outcome measurements|and statistical evaluation|devised))|((based on|used|determined|using|estimated|calculated).(a)?(power (calculation|analysis|estimat)))|(power (calculation|analysis|estimation|of at least|of the study was|of .?[0-9]{1,3}(.[0-9]{1,3})?%?)|(?<!not)adequate to detect|(to detect (a |the)?(differences? (of)?|treatment effect|treatment interaction|statistical differences|expected difference|predetermined effect|mean difference|(.?[0-

	9]{1,3}(.[0-9]{1,3})?%?) (improvement|increase|decrease)|similar treatment effect|significant change))|to insure sufficient power|would be needed if the null hypothesis|(to estimate the|used to determine the|based on these assumptions a) sample size|(the study|gave appropriate|increase the|descrease the|%|the planned|statistical) power|a power of|(?<!failed to)achieve statistical significance|was powered at|minimum number of (mice|rats|animals|subjects|patients) were used|% chance of|% to detect a|minimum clinically worthwhile effect|power of more than|% to reject the null hypothesis|effectively powered|power and statistical analysis|are required per group|per group were required|minimum number required to give|required to give statistically valid results|through a priori calculation)
COI	(author(s'|s)?.(adher|agree).(to |with)?(the)?.{0,20}(guide|polic))|(conflict(s)? of|competing) (financial)?interest|conflict(s)? declare|financial (conflict(s)?)?interest|inventor(s)? of the intellectual property
CAW	(treated|used|conducted|performed) in (accordance|adherence) with|Care and Use of Laboratory Animals|Animal Care(and Us(e|age))? (Committee|Guidelines)|Using Animals in Intramural Research|Animal Protection Bill|Bioethical Committee|Use and Care of Animals|(Council|Committee) (for|on) Animal Care|regulations for animal experimentation|ethical use of animals|(protocols|procedures|experiments|studies|work) (were|was|are) approved by|efforts were made to minimize the (number of animals and their)?suffering|Ethics Committee|Laboratory Animal (Welfare|Care)|in compliance with|International Council for Laboratory Animals|Animal Research Ethics

2. ***RegexROB R package (Bahor., Z., Hair, K., Liao, J., 2021)***

https://github.com/camaradesuk/RegexRoB/

3. ***Updated AutoAnnotation package (Bahor., Z., Hair, K., Liao, J., 2021)***

https://github.com/camaradesuk/AutoAnnotation/

Appendix C4

1. ***Protocol for evaluating automated deduplication tools***

https://doi.org/10.17605/OSF.IO/W3MAK

2. ***ASySD R package (Hair, K., 2021)***

https://github.com/camaradesuk/ASySD

3. ***ASySD Shiny application***

https://camarades.shinyapps.io/RDedup/

4. ***ASySD Shiny app R code (Hair, K., 2020)***

https://github.com/camaradesuk/ASySD_shiny

5. ***ASySD performance evaluation raw datasets and code***

https://doi.org/10.17605/OSF.IO/W3MAK

6. ***ASySD preprint:***

Hair, K., Bahor, Z., Macleod, M., Liao, J., & Sena, E. S. (2021). The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. https://doi.org/10.1101/2021.05.04.442412

Appendix C5

1. ***Synaptic plasticity and cognition in AD models SR protocol:***

https://doi.org /10.17605/OSF.IO/GFCVU

2. ***In vitro electrophysiology regex:***

[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

3. ***MWM regex:***

[Mm]orris [wW]ater|MWM|[Mm]orris [sS]wim|[Mm]orris [Ss]patial|(?<![Rr]adial|[Rr]adial [Aa]rm|[rR]adial-[aA]rm)[Ww]ater [mM]aze

4. ***Data extraction guide for reviewers***

https://osf.io/eayd7/

5. ***Publications included in SR***

Publications are listed with study ID used throughout review.

101. Chapman, P. F., White, G. L., Jones, M. W., Cooper-Blacketer, D., Marshall, V. J., Irizarry, M., . . . Hsiao, K. K. (1999). Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice. *Nat Neurosci, 2*(3), 271-276. doi:10.1038/6374
102. Jacobsen, J. S., Wu, C. C., Redwine, J. M., Comery, T. A., Arias, R., Bowlby, M., . . . Bloom, F. E. (2006). Early-onset behavioral and synaptic deficits in a mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A, 103*(13), 5161-5166. doi:10.1073/pnas.0600948103
103. Ma, H., Lesne, S., Kotilinek, L., Steidl-Nichols, J. V., Sherman, M., Younkin, L., . . . Ashe, K. H. (2007). Involvement of beta-site APP cleaving enzyme 1 (BACE1) in amyloid precursor protein-mediated enhancement of memory and activity-dependent synaptic plasticity. *Proc Natl Acad Sci U S A, 104*(19), 8167-8172. doi:10.1073/pnas.0609521104

104. Jacobsen, J. S., Comery, T. A., Martone, R. L., Elokdah, H., Crandall, D. L., Oganesian, A., . . . Pangalos, M. N. (2008). Enhanced clearance of Abeta in brain by sustaining the plasmin proteolysis cascade. *Proc Natl Acad Sci U S A, 105*(25), 8754-8759. doi:10.1073/pnas.0710823105
105. Hermann, D., Both, M., Ebert, U., Gross, G., Schoemaker, H., Draguhn, A., . . . Nimmrich, V. (2009). Synaptic transmission is impaired prior to plaque formation in amyloid precursor protein-overexpressing mice without altering behaviorally-correlated sharp wave-ripple complexes. *Neuroscience, 162*(4), 1081-1090. doi:10.1016/j.neuroscience.2009.05.044
106. Kiyota, T., Yamamoto, M., Xiong, H., Lambert, M. P., Klein, W. L., Gendelman, H. E., . . . Ikezu, T. (2009). CCL2 accelerates microglia-mediated Abeta oligomer formation and progression of neurocognitive dysfunction. *PLoS ONE, 4*(7), e6197. doi:10.1371/journal.pone.0006197
107. Gong, B., Chen, F., Pan, Y., Arrieta-Cruz, I., Yoshida, Y., Haroutunian, V., & Pasinetti, G. M. (2010). SCFFbx2-E3-ligase-mediated degradation of BACE1 attenuates Alzheimer's disease amyloidosis and improves synaptic function. *Aging Cell, 9*(6), 1018-1031. doi:10.1111/j.1474-9726.2010.00632.x
108. Ma, T., Hoeffer, C. A., Capetillo-Zarate, E., Yu, F., Wong, H., Lin, M. T., . . . Gouras, G. K. (2010). Dysregulation of the mTOR pathway mediates impairment of synaptic plasticity in a mouse model of Alzheimer's disease. *PLoS ONE, 5*(9). doi:10.1371/journal.pone.0012845
109. Townsend, M., Qu, Y., Gray, A., Wu, Z., Seto, T., Hutton, M., . . . Middleton, R. E. (2010). Oral treatment with a gamma-secretase inhibitor improves long-term potentiation in a mouse model of Alzheimer's disease. *J Pharmacol Exp Ther, 333*(1), 110-119. doi:10.1124/jpet.109.163691
110. Witton, J., Brown, J. T., Jones, M. W., & Randall, A. D. (2010). Altered synaptic plasticity in the mossy fibre pathway of transgenic mice expressing mutant amyloid precursor protein. *Mol Brain, 3*, 32. doi:10.1186/1756-6606-3-32
111. Balducci, C., Mehdawy, B., Mare, L., Giuliani, A., Lorenzini, L., Sivilia, S., . . . Nistico, R. (2011). The gamma-secretase modulator CHF5074 restores memory and hippocampal synaptic plasticity in plaque-free Tg2576 mice. *J Alzheimers Dis, 24*(4), 799-816. doi:10.3233/JAD-2011-101839
112. D'Amelio, M., Cavallucci, V., Middei, S., Marchetti, C., Pacioni, S., Ferri, A., . . . Cecconi, F. (2011). Caspase-3 triggers early synaptic dysfunction in a mouse model of Alzheimer's disease. *Nat Neurosci, 14*(1), 69-76. doi:10.1038/nn.2709
113. Jung, J. H., An, K., Kwon, O. B., Kim, H. S., & Kim, J. H. (2011). Pathway-specific alteration of synaptic plasticity in Tg2576 mice. *Mol Cells, 32*(2), 197-201. doi:10.1007/s10059-011-0077-8
114. Wen, L., Tang, F. L., Hong, Y., Luo, S. W., Wang, C. L., He, W., . . . Xiong, W. C. (2011). VPS35 haploinsufficiency increases Alzheimer's disease neuropathology. *J Cell Biol, 195*(5), 765-779. doi:10.1083/jcb.201105109
115. Wang, J., Ferruzzi, M. G., Ho, L., Blount, J., Janle, E. M., Gong, B., . . . Pasinetti, G. M. (2012). Brain-targeted proanthocyanidin metabolites for Alzheimer's disease treatment. *J Neurosci, 32*(15), 5144-5150. doi:10.1523/JNEUROSCI.6437-11.2012
116. Cavallucci, V., Berretta, N., Nobili, A., Nistico, R., Mercuri, N. B., & D'Amelio, M. (2013). Calcineurin inhibition rescues early synaptic plasticity deficits in a mouse model

of Alzheimer's disease. *Neuromolecular Med, 15*(3), 541-548. doi:10.1007/s12017-013-8241-2
117. Gong, B., Pan, Y., Zhao, W., Knable, L., Vempati, P., Begum, S., . . . Pasinetti, G. M. (2013). IVIG immunotherapy protects against synaptic dysfunction in Alzheimer's disease through complement anaphylatoxin C5a-mediated AMPA-CREB-C/EBP signaling pathway. *Mol Immunol, 56*(4), 619-629. doi:10.1016/j.molimm.2013.06.016
118. Gong, B., Pan, Y., Vempati, P., Zhao, W., Knable, L., Ho, L., . . . Pasinetti, G. M. (2013). Nicotinamide riboside restores cognition through an upregulation of proliferator-activated receptor-gamma coactivator 1alpha regulated beta-secretase 1 degradation and mitochondrial gene expression in Alzheimer's mouse models. *Neurobiol Aging, 34*(6), 1581-1588. doi:10.1016/j.neurobiolaging.2012.12.005
119. Ho, L., Ferruzzi, M. G., Janle, E. M., Wang, J., Gong, B., Chen, T. Y., . . . Pasinetti, G. M. (2013). Identification of brain-targeted bioactive dietary quercetin-3-O-glucuronide as a novel intervention for Alzheimer's disease. *FASEB J, 27*(2), 769-781. doi:10.1096/fj.12-212118
120. An, K., Jung, J. H., Jeong, A. Y., Kim, H. G., Jung, S. Y., Lee, K., . . . Kim, J. H. (2014). Neuritin can normalize neural deficits of Alzheimer's disease. *Cell Death Dis, 5*, e1523. doi:10.1038/cddis.2014.478
121. Lawrence, J. L., Tong, M., Alfulaij, N., Sherrin, T., Contarino, M., White, M. M., . . . Nichols, R. A. (2014). Regulation of presynaptic Ca2+, synaptic plasticity and contextual fear conditioning by a N-terminal beta-amyloid fragment. *J Neurosci, 34*(43), 14210-14218. doi:10.1523/JNEUROSCI.0326-14.2014
122. Lee, L., Dale, E., Staniszewski, A., Zhang, H., Saeed, F., Sakurai, M., . . . Arancio, O. (2014). Regulation of synaptic plasticity and cognition by SUMO in normal physiology and Alzheimer's disease. *Sci Rep, 4*, 7190. doi:10.1038/srep07190
123. Jiang, X., Chai, G. S., Wang, Z. H., Hu, Y., Li, X. G., Ma, Z. W., . . . Liu, G. P. (2015). Spatial training preserves associative memory capacity with augmentation of dendrite ramification and spine generation in Tg2576 mice. *Sci Rep, 5*, 9488. doi:10.1038/srep09488
124. Lante, F., Chafai, M., Raymond, E. F., Pereira, A. R., Mouska, X., Kootar, S., . . . Marie, H. (2015). Subchronic glucocorticoid receptor inhibition rescues early episodic memory and synaptic plasticity deficits in a mouse model of Alzheimer's disease. *Neuropsychopharmacology, 40*(7), 1772-1781. doi:10.1038/npp.2015.25
125. Murphy, M., Wilson, Y. M., Vargas, E., Munro, K. M., Smith, B., Huang, A., . . . Barrett, G. L. (2015). Reduction of p75 neurotrophin receptor ameliorates the cognitive deficits in a model of Alzheimer's disease. *Neurobiol Aging, 36*(2), 740-752. doi:10.1016/j.neurobiolaging.2014.09.014
126. Severini, C., Lattanzi, R., Maftei, D., Marconi, V., Ciotti, M. T., Petrocchi Passeri, P., . . . Negri, L. (2015). Bv8/prokineticin 2 is involved in Abeta-induced neurotoxicity. *Sci Rep, 5*, 15301. doi:10.1038/srep15301
127. Fernandez-Fernandez, D., Dorner-Ciossek, C., Kroker, K. S., & Rosenbrock, H. (2016). Age-related synaptic dysfunction in Tg2576 mice starts as a failure in early long-term potentiation which develops into a full abolishment of late long-term potentiation. *J Neurosci Res, 94*(3), 266-281. doi:10.1002/jnr.23701
128. Huh, S., Baek, S. J., Lee, K. H., Whitcomb, D. J., Jo, J., Choi, S. M., . . . Kim, B. C. (2016). The reemergence of long-term potentiation in aged Alzheimer's disease mouse model. *Sci Rep, 6*, 29152. doi:10.1038/srep29152

129. Shah, D., Praet, J., Latif Hernandez, A., Hofling, C., Anckaerts, C., Bard, F., . . . Van der Linden, A. (2016). Early pathologic amyloid induces hypersynchrony of BOLD resting-state networks in transgenic mice and provides an early therapeutic window before amyloid plaque deposition. *Alzheimers Dement, 12*(9), 964-976. doi:10.1016/j.jalz.2016.03.010
130. Nobili, A., Latagliata, E. C., Viscomi, M. T., Cavallucci, V., Cutuli, D., Giacovazzo, G., . . . D'Amelio, M. (2017). Dopamine neuronal loss contributes to memory and reward dysfunction in a model of Alzheimer's disease. *Nat Commun, 8*, 14727. doi:10.1038/ncomms14727
131. Ricciarelli, R., Brullo, C., Prickaerts, J., Arancio, O., Villa, C., Rebosio, C., . . . Fedele, E. (2017). Memory-enhancing effects of GEBR-32a, a new PDE4D inhibitor holding promise for the treatment of Alzheimer's disease. *Sci Rep, 7*, 46320. doi:10.1038/srep46320
132. Wang, X., Liu, D., Huang, H. Z., Wang, Z. H., Hou, T. Y., Yang, X., . . . Zhu, L. Q. (2018). A Novel MicroRNA-124/PTPN1 Signal Pathway Mediates Synaptic and Memory Deficits in Alzheimer's Disease. *Biol Psychiatry, 83*(5), 395-405. doi:10.1016/j.biopsych.2017.07.023
133. Arancio, O., Zhang, H. P., Chen, X., Lin, C., Trinchese, F., Puzzo, D., . . . Du Yan, S. S. (2004). RAGE potentiates Abeta-induced perturbation of neuronal function in transgenic mice. *EMBO J, 23*(20), 4096-4105. doi:10.1038/sj.emboj.7600415
134. Saura, C. A., Chen, G., Malkani, S., Choi, S. Y., Takahashi, R. H., Zhang, D., . . . Shen, J. (2005). Conditional inactivation of presenilin 1 prevents amyloid accumulation and temporarily rescues contextual and spatial working memory impairments in amyloid precursor protein transgenic mice. *J Neurosci, 25*(29), 6755-6764. doi:10.1523/JNEUROSCI.1247-05.2005
135. Palop, J. J., Chin, J., Roberson, E. D., Wang, J., Thwin, M. T., Bien-Ly, N., . . . Mucke, L. (2007). Aberrant excitatory neuronal activity and compensatory remodeling of inhibitory hippocampal circuits in mouse models of Alzheimer's disease. *Neuron, 55*(5), 697-711. doi:10.1016/j.neuron.2007.07.025
136. Du, H., Guo, L., Fang, F., Chen, D., Sosunov, A. A., McKhann, G. M., . . . Yan, S. D. (2008). Cyclophilin D deficiency attenuates mitochondrial and neuronal perturbation and ameliorates learning and memory in Alzheimer's disease. *Nat Med, 14*(10), 1097-1105. doi:10.1038/nm.1868
137. Sun, B., Zhou, Y., Halabisky, B., Lo, I., Cho, S. H., Mueller-Steiner, S., . . . Gan, L. (2008). Cystatin C-cathepsin B axis regulates amyloid beta levels and associated neuronal deficits in an animal model of Alzheimer's disease. *Neuron, 60*(2), 247-257. doi:10.1016/j.neuron.2008.10.001
138. Harris, J. A., Devidze, N., Halabisky, B., Lo, I., Thwin, M. T., Yu, G. Q., . . . Mucke, L. (2010). Many neuronal and behavioral impairments in transgenic mouse models of Alzheimer's disease are independent of caspase cleavage of the amyloid precursor protein. *J Neurosci, 30*(1), 372-381. doi:10.1523/JNEUROSCI.5341-09.2010
139. Cisse, M., Halabisky, B., Harris, J., Devidze, N., Dubal, D. B., Sun, B., . . . Mucke, L. (2011). Reversing EphB2 depletion rescues cognitive functions in Alzheimer model. *Nature, 469*(7328), 47-52. doi:10.1038/nature09635
140. Kiyota, T., Ingraham, K. L., Jacobsen, M. T., Xiong, H., & Ikezu, T. (2011). FGF2 gene transfer restores hippocampal functions in mouse models of Alzheimer's disease and

has therapeutic implications for neurocognitive disorders. *Proc Natl Acad Sci U S A, 108*(49), E1339-1348. doi:10.1073/pnas.1102349108
141. Sanchez, P. E., Zhu, L., Verret, L., Vossel, K. A., Orr, A. G., Cirrito, J. R., . . . Mucke, L. (2012). Levetiracetam suppresses neuronal network dysfunction and reverses synaptic and cognitive deficits in an Alzheimer's disease model. *Proc Natl Acad Sci U S A, 109*(42), E2895-2903. doi:10.1073/pnas.1121081109
142. Dubal, D. B., Zhu, L., Sanchez, P. E., Worden, K., Broestl, L., Johnson, E., . . . Mucke, L. (2015). Life extension factor klotho prevents mortality and enhances cognition in hAPP transgenic mice. *J Neurosci, 35*(6), 2358-2371. doi:10.1523/JNEUROSCI.5791-12.2015
143. Fang, D., Wang, Y., Zhang, Z., Du, H., Yan, S., Sun, Q., . . . Yan, S. S. (2015). Increased neuronal PreP activity reduces Abeta accumulation, attenuates neuroinflammation and improves mitochondrial and synaptic function in Alzheimer disease's mouse model. *Hum Mol Genet, 24*(18), 5198-5210. doi:10.1093/hmg/ddv241
144. Letronne, F., Laumet, G., Ayral, A. M., Chapuis, J., Demiautte, F., Laga, M., . . . Lambert, J. C. (2016). ADAM30 Downregulates APP-Linked Defects Through Cathepsin D Activation in Alzheimer's Disease. *EBioMedicine, 9*, 278-292. doi:10.1016/j.ebiom.2016.06.002
145. Qu, X., Yuan, F. N., Corona, C., Pasini, S., Pero, M. E., Gundersen, G. G., . . . Bartolini, F. (2017). Stabilization of dynamic microtubules by mDia1 drives Tau-dependent Abeta1-42 synaptotoxicity. *J Cell Biol, 216*(10), 3161-3178. doi:10.1083/jcb.201701045
146. Zhang, H., Zhang, L., Zhou, D., He, X., Wang, D., Pan, H., . . . Sun, B. (2017). Ablating ErbB4 in PV neurons attenuates synaptic and cognitive deficits in an animal model of Alzheimer's disease. *Neurobiol Dis, 106*, 171-180. doi:10.1016/j.nbd.2017.07.001
147. Kimura, R., & Ohno, M. (2009). Impairments in remote memory stabilization precede hippocampal synaptic and cognitive failures in 5XFAD Alzheimer mouse model. *Neurobiol Dis, 33*(2), 229-235. doi:10.1016/j.nbd.2008.10.006
148. Kimura, R., Devi, L., & Ohno, M. (2010). Partial reduction of BACE1 improves synaptic plasticity, recent and remote memories in Alzheimer's disease transgenic mice. *J Neurochem, 113*(1), 248-261. doi:10.1111/j.1471-4159.2010.06608.x
149. Chen, R., Zhang, J., Wu, Y., Wang, D., Feng, G., Tang, Y. P., . . . Chen, C. (2012). Monoacylglycerol lipase is a therapeutic target for Alzheimer's disease. *Cell Rep, 2*(5), 1329-1339. doi:10.1016/j.celrep.2012.09.030
150. Crouzin, N., Baranger, K., Cavalier, M., Marchalant, Y., Cohen-Solal, C., Roman, F. S., . . . Vignes, M. (2013). Area-specific alterations of synaptic plasticity in the 5XFAD mouse model of Alzheimer's disease: dissociation between somatosensory cortex and hippocampus. *PLoS ONE, 8*(9), e74667. doi:10.1371/journal.pone.0074667
151. Seo, J., Giusti-Rodriguez, P., Zhou, Y., Rudenko, A., Cho, S., Ota, K. T., . . . Tsai, L. H. (2014). Activity-dependent p25 generation regulates synaptic plasticity and Abeta-induced cognitive impairment. *Cell, 157*(2), 486-498. doi:10.1016/j.cell.2014.01.065
152. Wu, Z., Guo, Z., Gearing, M., & Chen, G. (2014). Tonic inhibition in dentate gyrus impairs long-term potentiation and memory in an Alzheimer's [corrected] disease model. *Nat Commun, 5*, 4159. doi:10.1038/ncomms5159
153. Zhang, J., Hu, M., Teng, Z., Tang, Y. P., & Chen, C. (2014). Synaptic and cognitive improvements by inhibition of 2-AG metabolism are through upregulation of microRNA-188-3p in a mouse model of Alzheimer's disease. *J Neurosci, 34*(45), 14919-14933. doi:10.1523/JNEUROSCI.1165-14.2014

154. Zhang, Z., Song, M., Liu, X., Su Kang, S., Duong, D. M., Seyfried, N. T., . . . Ye, K. (2015). Delta-secretase cleaves amyloid precursor protein and regulates the pathogenesis in Alzheimer's disease. *Nat Commun, 6*, 8762. doi:10.1038/ncomms9762
155. Baranger, K., Marchalant, Y., Bonnet, A. E., Crouzin, N., Carrete, A., Paumier, J. M., . . . Rivera, S. (2016). MT5-MMP is a new pro-amyloidogenic proteinase that promotes amyloid pathology and cognitive decline in a transgenic mouse model of Alzheimer's disease. *Cell Mol Life Sci, 73*(1), 217-236. doi:10.1007/s00018-015-1992-1
156. Lee, K., Kim, H., An, K., Kwon, O. B., Park, S., Cha, J. H., . . . Kim, H. S. (2016). Replenishment of microRNA-188-5p restores the synaptic and cognitive deficits in 5XFAD Mouse Model of Alzheimer's Disease. *Sci Rep, 6*, 34433. doi:10.1038/srep34433
157. Colie, S., Sarroca, S., Palenzuela, R., Garcia, I., Matheu, A., Corpas, R., . . . Nebreda, A. R. (2017). Neuronal p38alpha mediates synaptic and cognitive dysfunction in an Alzheimer's mouse model by controlling beta-amyloid production. *Sci Rep, 7*, 45306. doi:10.1038/srep45306
158. Duran-Aniotz, C., Cornejo, V. H., Espinoza, S., Ardiles, A. O., Medinas, D. B., Salazar, C., . . . Hetz, C. (2017). IRE1 signaling exacerbates Alzheimer's disease pathogenesis. *Acta Neuropathol, 134*(3), 489-506. doi:10.1007/s00401-017-1694-x
159. Hwang, K. D., Bak, M. S., Kim, S. J., Rhee, S., & Lee, Y. S. (2017). Restoring synaptic plasticity and memory in mouse models of Alzheimer's disease by PKR inhibition. *Mol Brain, 10*(1), 57. doi:10.1186/s13041-017-0338-3
160. MacPherson, K. P., Sompol, P., Kannarkat, G. T., Chang, J., Sniffen, L., Wildner, M. E., . . . Tansey, M. G. (2017). Peripheral administration of the soluble TNF inhibitor XPro1595 modifies brain immune cell profiles, decreases beta-amyloid plaque load, and rescues impaired long-term potentiation in 5xFAD mice. *Neurobiol Dis, 102*, 81-95. doi:10.1016/j.nbd.2017.02.010
161. Maezawa, I., Zou, B., Di Lucente, J., Cao, W. S., Pascual, C., Weerasekara, S., . . . Jin, L. W. (2017). The Anti-Amyloid-beta and Neuroprotective Properties of a Novel Tricyclic Pyrone Molecule. *J Alzheimers Dis, 58*(2), 559-574. doi:10.3233/JAD-161175
162. Zhang, Z., Obianyo, O., Dall, E., Du, Y., Fu, H., Liu, X., . . . Ye, K. (2017). Inhibition of delta-secretase improves cognitive functions in mouse models of Alzheimer's disease. *Nat Commun, 8*, 14740. doi:10.1038/ncomms14740
163. Zhen, J., Qian, Y., Weng, X., Su, W., Zhang, J., Cai, L., . . . Wang, X. (2017). Gamma rhythm low field magnetic stimulation alleviates neuropathologic changes and rescues memory and cognitive impairments in a mouse model of Alzheimer's disease. *Alzheimers Dement (N Y), 3*(4), 487-497. doi:10.1016/j.trci.2017.07.002
164. Trinchese, F., Liu, S., Battaglia, F., Walter, S., Mathews, P. M., & Arancio, O. (2004). Progressive age-related development of Alzheimer-like pathology in APP/PS1 mice. *Ann Neurol, 55*(6), 801-814. doi:10.1002/ana.20101
165. Trinchese, F., Fa, M., Liu, S., Zhang, H., Hidalgo, A., Schmidt, S. D., . . . Arancio, O. (2008). Inhibition of calpains improves memory and synaptic transmission in a mouse model of Alzheimer disease. *J Clin Invest, 118*(8), 2796-2807. doi:10.1172/JCI34254
166. Yoshiike, Y., Kimura, T., Yamashita, S., Furudate, H., Mizoroki, T., Murayama, M., & Takashima, A. (2008). GABA(A) receptor-mediated acceleration of aging-associated memory decline in APP/PS1 mice and its pharmacological treatment by picrotoxin. *PLoS ONE, 3*(8), e3029. doi:10.1371/journal.pone.0003029
167. Puzzo, D., Staniszewski, A., Deng, S. X., Privitera, L., Leznik, E., Liu, S., . . . Arancio, O. (2009). Phosphodiesterase 5 inhibition improves synaptic function, memory, and

amyloid-beta load in an Alzheimer's disease mouse model. *J Neurosci, 29*(25), 8075-8086. doi:10.1523/JNEUROSCI.0864-09.2009
168. Gimbel, D. A., Nygaard, H. B., Coffey, E. E., Gunther, E. C., Lauren, J., Gimbel, Z. A., & Strittmatter, S. M. (2010). Memory impairment in transgenic Alzheimer mice requires cellular prion protein. *J Neurosci, 30*(18), 6367-6374. doi:10.1523/JNEUROSCI.0395-10.2010
169. Volianskis, A., Kostner, R., Molgaard, M., Hass, S., & Jensen, M. S. (2010). Episodic memory deficits are not related to altered glutamatergic synaptic transmission and plasticity in the CA1 hippocampus of the APPswe/PS1deltaE9-deleted transgenic mice model of ss-amyloidosis. *Neurobiol Aging, 31*(7), 1173-1187. doi:10.1016/j.neurobiolaging.2008.08.005
170. Hsiao, Y. H., Chen, P. S., Chen, S. H., & Gean, P. W. (2011). The involvement of Cdk5 activator p35 in social isolation-triggered onset of early Alzheimer's disease-related cognitive deficit in the transgenic mice. *Neuropsychopharmacology, 36*(9), 1848-1858. doi:10.1038/npp.2011.69
171. Inestrosa, N. C., Tapia-Rojas, C., Griffith, T. N., Carvajal, F. J., Benito, M. J., Rivera-Dictter, A., . . . Varela-Nallar, L. (2011). Tetrahydrohyperforin prevents cognitive deficit, Abeta deposition, tau phosphorylation and synaptotoxicity in the APPswe/PSEN1DeltaE9 model of Alzheimer's disease: a possible effect on APP processing. *Transl Psychiatry, 1*, e20. doi:10.1038/tp.2011.19
172. Ma, T., Hoeffer, C. A., Wong, H., Massaad, C. A., Zhou, P., Iadecola, C., . . . Klann, E. (2011). Amyloid beta-induced impairments in hippocampal synaptic plasticity are rescued by decreasing mitochondrial superoxide. *J Neurosci, 31*(15), 5589-5595. doi:10.1523/JNEUROSCI.6566-10.2011
173. Soderman, A., Mikkelsen, J. D., West, M. J., Christensen, D. Z., & Jensen, M. S. (2011). Activation of nicotinic alpha(7) acetylcholine receptor enhances long term potentation in wild type mice but not in APP(swe)/PS1DeltaE9 mice. *Neurosci Lett, 487*(3), 325-329. doi:10.1016/j.neulet.2010.10.049
174. Furman, J. L., Sama, D. M., Gant, J. C., Beckett, T. L., Murphy, M. P., Bachstetter, A. D., . . . Norris, C. M. (2012). Targeting astrocytes ameliorates neurologic changes in a mouse model of Alzheimer's disease. *J Neurosci, 32*(46), 16129-16140. doi:10.1523/JNEUROSCI.2323-12.2012
175. Heneka, M. T., Kummer, M. P., Stutz, A., Delekate, A., Schwartz, S., Vieira-Saecker, A., . . . Golenbock, D. T. (2013). NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature, 493*(7434), 674-678. doi:10.1038/nature11729
176. Inestrosa, N. C., Carvajal, F. J., Zolezzi, J. M., Tapia-Rojas, C., Serrano, F., Karmelic, D., . . . Santos, M. J. (2013). Peroxisome proliferators reduce spatial memory impairment, synaptic failure, and neurodegeneration in brains of a double transgenic mice model of Alzheimer's disease. *J Alzheimers Dis, 33*(4), 941-959. doi:10.3233/JAD-2012-120397
177. Kelly, R. J., Minogue, A. M., Lyons, A., Jones, R. S., Browne, T. C., Costello, D. A., . . . Lynch, M. A. (2013). Glial Activation in AbetaPP/PS1 Mice is Associated with Infiltration of IFNgamma-Producing Cells. *J Alzheimers Dis, 37*(1), 63-75. doi:10.3233/JAD-130539
178. Ma, T., Trinh, M. A., Wexler, A. J., Bourbon, C., Gatti, E., Pierre, P., . . . Klann, E. (2013). Suppression of eIF2alpha kinases alleviates Alzheimer's disease-related plasticity and memory deficits. *Nat Neurosci, 16*(9), 1299-1305. doi:10.1038/nn.3486

179. Toth, M. E., Szegedi, V., Varga, E., Juhasz, G., Horvath, J., Borbely, E., . . . Santha, M. (2013). Overexpression of Hsp27 ameliorates symptoms of Alzheimer's disease in APP/PS1 mice. *Cell Stress Chaperones, 18*(6), 759-771. doi:10.1007/s12192-013-0428-9
180. Fu, A. K., Hung, K. W., Huang, H., Gu, S., Shen, Y., Cheng, E. Y., . . . Ip, N. Y. (2014). Blockade of EphA4 signaling ameliorates hippocampal synaptic dysfunctions in mouse models of Alzheimer's disease. *Proc Natl Acad Sci U S A, 111*(27), 9959-9964. doi:10.1073/pnas.1405803111
181. Hong, X., Liu, J., Zhu, G., Zhuang, Y., Suo, H., Wang, P., . . . Huang, F. (2014). Parkin overexpression ameliorates hippocampal long-term potentiation and beta-amyloid load in an Alzheimer's disease mouse model. *Hum Mol Genet, 23*(4), 1056-1072. doi:10.1093/hmg/ddt501
182. Kummer, M. P., Hammerschmidt, T., Martinez, A., Terwel, D., Eichele, G., Witten, A., . . . Urban, I. (2014). Ear2 deletion causes early memory and learning deficits in APP/PS1 mice. *J Neurosci, 34*(26), 8845-8854. doi:10.1523/JNEUROSCI.4027-13.2014
183. Ma, T., Chen, Y., Vingtdeux, V., Zhao, H., Viollet, B., Marambaud, P., & Klann, E. (2014). Inhibition of AMP-activated protein kinase signaling alleviates impairments in hippocampal synaptic plasticity induced by amyloid beta. *J Neurosci, 34*(36), 12230-12238. doi:10.1523/JNEUROSCI.1694-14.2014
184. Metais, C., Brennan, K., Mably, A. J., Scott, M., Walsh, D. M., & Herron, C. E. (2014). Simvastatin treatment preserves synaptic plasticity in AbetaPPswe/PS1dE9 mice. *J Alzheimers Dis, 39*(2), 315-329. doi:10.3233/JAD-130257
185. Serrano, F. G., Tapia-Rojas, C., Carvajal, F. J., Hancke, J., Cerpa, W., & Inestrosa, N. C. (2014). Andrographolide reduces cognitive impairment in young and mature AbetaPPswe/PS-1 mice. *Mol Neurodegener, 9*, 61. doi:10.1186/1750-1326-9-61
186. Vargas, J. Y., Fuenzalida, M., & Inestrosa, N. C. (2014). In vivo activation of Wnt signaling pathway enhances cognitive function of adult mice and reverses cognitive deficits in an Alzheimer's disease model. *J Neurosci, 34*(6), 2191-2202. doi:10.1523/JNEUROSCI.0862-13.2014
187. Vegh, M. J., Heldring, C. M., Kamphuis, W., Hijazi, S., Timmerman, A. J., Li, K. W., . . . van Kesteren, R. E. (2014). Reducing hippocampal extracellular matrix reverses early memory deficits in a mouse model of Alzheimer's disease. *Acta Neuropathol Commun, 2*, 76. doi:10.1186/s40478-014-0076-z
188. Zhang, W., Wang, P. J., Sha, H. Y., Ni, J., Li, M. H., & Gu, G. J. (2014). Neural stem cell transplants improve cognitive function without altering amyloid pathology in an APP/PS1 double transgenic model of Alzheimer's disease. *Mol Neurobiol, 50*(2), 423-437. doi:10.1007/s12035-014-8640-x
189. Zhang, M. Y., Zheng, C. Y., Zou, M. M., Zhu, J. W., Zhang, Y., Wang, J., . . . Xu, R. X. (2014). Lamotrigine attenuates deficits in synaptic plasticity and accumulation of amyloid plaques in APP/PS1 transgenic mice. *Neurobiol Aging, 35*(12), 2713-2725. doi:10.1016/j.neurobiolaging.2014.06.009
190. Zhao, L., Gong, N., Liu, M., Pan, X., Sang, S., Sun, X., . . . Xu, T. (2014). Beneficial synergistic effects of microdose lithium with pyrroloquinoline quinone in an Alzheimer's disease mouse model. *Neurobiol Aging, 35*(12), 2736-2745. doi:10.1016/j.neurobiolaging.2014.06.003
191. Cisternas, P., Lindsay, C. B., Salazar, P., Silva-Alvarez, C., Retamales, R. M., Serrano, F. G., . . . Inestrosa, N. C. (2015). The increased potassium intake improves cognitive

performance and attenuates histopathological markers in a model of Alzheimer's disease. *Biochim Biophys Acta, 1852*(12), 2630-2644. doi:10.1016/j.bbadis.2015.09.009
192. Hu, W. Y., He, Z. Y., Yang, L. J., Zhang, M., Xing, D., & Xiao, Z. C. (2015). The Ca(2+) channel inhibitor 2-APB reverses beta-amyloid-induced LTP deficit in hippocampus by blocking BAX and caspase-3 hyperactivation. *Br J Pharmacol, 172*(9), 2273-2285. doi:10.1111/bph.13048
193. Kim, H. Y., Kim, H. V., Jo, S., Lee, C. J., Choi, S. Y., Kim, D. J., & Kim, Y. (2015). EPPS rescues hippocampus-dependent cognitive deficits in APP/PS1 mice by disaggregation of amyloid-beta oligomers and plaques. *Nat Commun, 6*, 8997. doi:10.1038/ncomms9997
194. Megill, A., Tran, T., Eldred, K., Lee, N. J., Wong, P. C., Hoe, H. S., . . . Lee, H. K. (2015). Defective Age-Dependent Metaplasticity in a Mouse Model of Alzheimer's Disease. *J Neurosci, 35*(32), 11346-11357. doi:10.1523/JNEUROSCI.5289-14.2015
195. Woo, J. A., Boggess, T., Uhlar, C., Wang, X., Khan, H., Cappos, G., . . . Kang, D. E. (2015). RanBP9 at the intersection between cofilin and Abeta pathologies: rescue of neurodegenerative changes by RanBP9 reduction. *Cell Death Dis, 6*, 1676. doi:10.1038/cddis.2015.37
196. Woo, J. A., Zhao, X., Khan, H., Penn, C., Wang, X., Joly-Amado, A., . . . Kang, D. E. (2015). Slingshot-Cofilin activation mediates mitochondrial and synaptic dysfunction via Abeta ligation to beta1-integrin conformers. *Cell Death Differ, 22*(6), 921-934. doi:10.1038/cdd.2015.5
197. Fu, A. K., Hung, K. W., Yuen, M. Y., Zhou, X., Mak, D. S., Chan, I. C., . . . Ip, N. Y. (2016). IL-33 ameliorates Alzheimer's disease-like pathology and cognitive decline. *Proc Natl Acad Sci U S A, 113*(19), E2705-2713. doi:10.1073/pnas.1604032113
198. Kajiwara, Y., McKenzie, A., Dorr, N., Gama Sosa, M. A., Elder, G., Schmeidler, J., . . . Buxbaum, J. D. (2016). The human-specific CASP4 gene product contributes to Alzheimer-related synaptic and behavioural deficits. *Hum Mol Genet, 25*(19), 4315-4327. doi:10.1093/hmg/ddw265
199. Li, F., Wu, X., Li, J., & Niu, Q. (2016). Ginsenoside Rg1 ameliorates hippocampal long-term potentiation and memory in an Alzheimer's disease model. *Mol Med Rep, 13*(6), 4904-4910. doi:10.3892/mmr.2016.5103
200. Montgomery, K. S., Edwards, G., 3rd, Levites, Y., Kumar, A., Myers, C. E., Gluck, M. A., . . . Bizon, J. L. (2016). Deficits in hippocampal-dependent transfer generalization learning accompany synaptic dysfunction in a mouse model of amyloidosis. *Hippocampus, 26*(4), 455-471. doi:10.1002/hipo.22535
201. H., O. T. D. B. A. V. d. W. I. V. H. K. H. L. S. H. K. J. A. a. D. W. (2016). Age-dependent concomitant changes in synaptic function and GABAergic pathway in the APP/PS1 mouse model. *76*(4), 282-293.
202. Shen, Y., Tian, M., Zheng, Y., Gong, F., Fu, A. K. Y., & Ip, N. Y. (2016). Stimulation of the Hippocampal POMC/MC4R Circuit Alleviates Synaptic Plasticity Impairment in an Alzheimer's Disease Model. *Cell Rep, 17*(7), 1819-1831. doi:10.1016/j.celrep.2016.10.043
203. Yang, W., Zhou, X., Zimmermann, H. R., Cavener, D. R., Klann, E., & Ma, T. (2016). Repression of the eIF2alpha kinase PERK alleviates mGluR-LTD impairments in a mouse model of Alzheimer's disease. *Neurobiol Aging, 41*, 19-24. doi:10.1016/j.neurobiolaging.2016.02.005

204. Yang, Y. J., Zhao, Y., Yu, B., Xu, G. G., Wang, W., Zhan, J. Q., . . . Wei, B. (2016). GluN2B-containing NMDA receptors contribute to the beneficial effects of hydrogen sulfide on cognitive and synaptic plasticity deficits in APP/PS1 transgenic mice. *Neuroscience, 335*, 170-183. doi:10.1016/j.neuroscience.2016.08.033
205. Alves, S., Churlaud, G., Audrain, M., Michaelsen-Preusse, K., Fol, R., Souchet, B., . . . Cartier, N. (2017). Interleukin-2 improves amyloid pathology, synaptic failure and memory in Alzheimer's disease mice. *Brain, 140*(3), 826-842. doi:10.1093/brain/aww330
206. Gomez-Gonzalo, M., Martin-Fernandez, M., Martinez-Murillo, R., Mederos, S., Hernandez-Vivanco, A., Jamison, S., . . . Araque, A. (2017). Neuron-astrocyte signaling is preserved in the aging brain. *GLIA, 65*(4), 569-580. doi:10.1002/glia.23112
207. Jin, W. S., Shen, L. L., Bu, X. L., Zhang, W. W., Chen, S. H., Huang, Z. L., . . . Wang, Y. J. (2017). Peritoneal dialysis reduces amyloid-beta plasma levels in humans and attenuates Alzheimer-associated phenotypes in an APP/PS1 mouse model. *Acta Neuropathol, 134*(2), 207-220. doi:10.1007/s00401-017-1721-y
208. Li, Q., Navakkode, S., Rothkegel, M., Soong, T. W., Sajikumar, S., & Korte, M. (2017). Metaplasticity mechanisms restore plasticity and associativity in an animal model of Alzheimer's disease. *Proc Natl Acad Sci U S A, 114*(21), 5527-5532. doi:10.1073/pnas.1613700114
209. Maezawa, I., Nguyen, H. M., Di Lucente, J., Jenkins, D. P., Singh, V., Hilt, S., . . . Jin, L. W. (2018). Kv1.3 inhibition as a potential microglia-targeted therapy for Alzheimer's disease: preclinical proof of concept. *Brain, 141*(2), 596-612. doi:10.1093/brain/awx346
210. Tao, C. C., Hsu, W. L., Ma, Y. L., Cheng, S. J., & Lee, E. H. (2017). Epigenetic regulation of HDAC1 SUMOylation as an endogenous neuroprotection against Abeta toxicity in a mouse model of Alzheimer's disease. *Cell Death Differ, 24*(4), 597-614. doi:10.1038/cdd.2016.161
211. Yu, L., Liu, Y., Yang, H., Zhu, X., Cao, X., Gao, J., . . . Xu, Y. (2017). PSD-93 Attenuates Amyloid-beta-Mediated Cognitive Dysfunction by Promoting the Catabolism of Amyloid-beta. *J Alzheimers Dis, 59*(3), 913-927. doi:10.3233/JAD-170320
212. Martinez Hernandez, A., Urbanke, H., Gillman, A. L., Lee, J., Ryazanov, S., Agbemenyah, H. Y., . . . Fischer, A. (2018). The diphenylpyrazole compound anle138b blocks Abeta channels and rescues disease phenotypes in a mouse model for amyloid pathology. *EMBO Mol Med, 10*(1), 32-47. doi:10.15252/emmm.201707825
213. Wang, S. S., Jia, J., & Wang, Z. (2018). Mesenchymal Stem Cell-Derived Extracellular Vesicles Suppresses iNOS Expression and Ameliorates Neural Impairment in Alzheimer's Disease Mice. *J Alzheimers Dis, 61*(3), 1005-1013. doi:10.3233/JAD-170848
214. Zhu, L., Yang, L., Zhao, X., Liu, D., Guo, X., Liu, P., . . . Zou, L. (2018). Xanthoceraside modulates NR2B-containing NMDA receptors at synapses and rescues learning-memory deficits in APP/PS1 transgenic mice. *Psychopharmacology (Berl), 235*(1), 337-349. doi:10.1007/s00213-017-4775-6
215. Gureviciene, I., Ikonen, S., Gurevicius, K., Sarkaki, A., van Groen, T., Pussinen, R., . . . Tanila, H. (2004). Normal induction but accelerated decay of LTP in APP + PS1 transgenic mice. *Neurobiol Dis, 15*(2), 188-195. doi:10.1016/j.nbd.2003.11.011
216. Fitzjohn, S. M., Kuenzi, F., Morton, R. A., Rosahl, T. W., Lewis, H., Smith, D., . . . Collingridge, G. L. (2010). A study of long-term potentiation in transgenic mice over-expressing mutant forms of both amyloid precursor protein and presenilin-1. *Mol Brain, 3*(1), 21. doi:10.1186/1756-6606-3-21

217. Ricoy, U. M., Mao, P., Manczak, M., Reddy, P. H., & Frerking, M. E. (2011). A transgenic mouse model for Alzheimer's disease has impaired synaptic gain but normal synaptic dynamics. *Neurosci Lett, 500*(3), 212-215. doi:10.1016/j.neulet.2011.06.043
218. Maingret, V., Barthet, G., Deforges, S., Jiang, N., Mulle, C., & Amedee, T. (2017). PGE2-EP3 signaling pathway impairs hippocampal presynaptic long-term plasticity in a mouse model of Alzheimer's disease. *Neurobiol Aging, 50*, 13-24. doi:10.1016/j.neurobiolaging.2016.10.012
219. Jolas, T., Zhang, X. S., Zhang, Q., Wong, G., Del Vecchio, R., Gold, L., & Priestley, T. (2002). Long-term potentiation is increased in the CA1 area of the hippocampus of APP(swe/ind) CRND8 mice. *Neurobiol Dis, 11*(3), 394-409. doi:10.1006/nbdi.2002.0557
220. Arrieta-Cruz, I., Pavlides, C., & Pasinetti, G. (2010). Deep brain stimulation facilitates memory in a model of Alzheimer's disease. *Translational Neuroscience, 1*(2), 170-176. doi:10.2478/v10134-010-0026-7
221. Arrieta-Cruz, I., Wang, J., Pavlides, C., & Pasinetti, G. M. (2010). Carvedilol reestablishes long-term potentiation in a mouse model of Alzheimer's disease. *J Alzheimers Dis, 21*(2), 649-654. doi:10.3233/JAD-2010-100225
222. Ye, H., Jalini, S., Mylvaganam, S., & Carlen, P. (2010). Activation of large-conductance Ca(2+)-activated K(+) channels depresses basal synaptic transmission in the hippocampal CA1 area in APP (swe/ind) TgCRND8 mice. *Neurobiol Aging, 31*(4), 591-604. doi:10.1016/j.neurobiolaging.2008.05.012
223. Sclip, A., Antoniou, X., Colombo, A., Camici, G. G., Pozzi, L., Cardinetti, D., . . . Borsello, T. (2011). c-Jun N-terminal kinase regulates soluble Abeta oligomers and cognitive impairment in AD mouse model. *J Biol Chem, 286*(51), 43871-43880. doi:10.1074/jbc.M111.297515
224. Sclip, A., Tozzi, A., Abaza, A., Cardinetti, D., Colombo, I., Calabresi, P., . . . Borsello, T. (2014). c-Jun N-terminal kinase has a key role in Alzheimer disease synaptic dysfunction in vivo. *Cell Death Dis, 5*, e1019. doi:10.1038/cddis.2013.559
225. Luccarini, I., Grossi, C., Rigacci, S., Coppi, E., Pugliese, A. M., Pantano, D., . . . Casamenti, F. (2015). Oleuropein aglycone protects against pyroglutamylated-3 amyloid-ss toxicity: biochemical, epigenetic and functional correlates. *Neurobiol Aging, 36*(2), 648-663. doi:10.1016/j.neurobiolaging.2014.08.029
226. Tozzi, A., Sclip, A., Tantucci, M., de Iure, A., Ghiglieri, V., Costa, C., . . . Calabresi, P. (2015). Region- and age-dependent reductions of hippocampal long-term potentiation and NMDA to AMPA ratio in a genetic model of Alzheimer's disease. *Neurobiol Aging, 36*(1), 123-133. doi:10.1016/j.neurobiolaging.2014.07.002
227. Cavanagh, C., Tse, Y. C., Nguyen, H. B., Krantic, S., Breitner, J. C., Quirion, R., & Wong, T. P. (2016). Inhibiting tumor necrosis factor-alpha before amyloidosis prevents synaptic deficits in an Alzheimer's disease model. *Neurobiol Aging, 47*, 41-49. doi:10.1016/j.neurobiolaging.2016.07.009
228. Hinrich, A. J., Jodelka, F. M., Chang, J. L., Brutman, D., Bruno, A. M., Briggs, C. A., . . . Hastings, M. L. (2016). Therapeutic correction of ApoER2 splicing in Alzheimer's disease mice using antisense oligonucleotides. *EMBO Mol Med, 8*(4), 328-345. doi:10.15252/emmm.201505846
229. Kimura, R., MacTavish, D., Yang, J., Westaway, D., & Jhamandas, J. H. (2017). Pramlintide Antagonizes Beta Amyloid (Abeta)- and Human Amylin-Induced Depression of Hippocampal Long-Term Potentiation. *Mol Neurobiol, 54*(1), 748-754. doi:10.1007/s12035-016-9684-x

230. Knock, E., Matsuzaki, S., Takamura, H., Satoh, K., Rooke, G., Han, K., . . . Fraser, P. E. (2018). SUMO1 impact on Alzheimer disease pathology in an amyloid-depositing mouse model. *Neurobiol Dis, 110*, 154-165. doi:10.1016/j.nbd.2017.11.015
231. Oddo, S., Caccamo, A., Shepherd, J. D., Murphy, M. P., Golde, T. E., Kayed, R., . . . LaFerla, F. M. (2003). Triple-Transgenic Model of Alzheimer's Disease with Plaques and Tangles. *Neuron, 39*(3), 409-421. doi:10.1016/s0896-6273(03)00434-3
232. Chakroborty, S., Goussakov, I., Miller, M. B., & Stutzmann, G. E. (2009). Deviant ryanodine receptor-mediated calcium release resets synaptic homeostasis in presymptomatic 3xTg-AD mice. *J Neurosci, 29*(30), 9458-9470. doi:10.1523/JNEUROSCI.2047-09.2009
233. Chakroborty, S., Kim, J., Schneider, C., Jacobson, C., Molgo, J., & Stutzmann, G. E. (2012). Early presynaptic and postsynaptic calcium signaling abnormalities mask underlying synaptic depression in presymptomatic Alzheimer's disease mice. *J Neurosci, 32*(24), 8341-8353. doi:10.1523/JNEUROSCI.0936-12.2012
234. Chakroborty, S., Briggs, C., Miller, M. B., Goussakov, I., Schneider, C., Kim, J., . . . Stutzmann, G. E. (2012). Stabilizing ER Ca2+ channel function as an early preventative strategy for Alzheimer's disease. *PLoS ONE, 7*(12), e52056. doi:10.1371/journal.pone.0052056
235. Hsiao, Y. H., Kuo, J. R., Chen, S. H., & Gean, P. W. (2012). Amelioration of social isolation-triggered onset of early Alzheimer's disease-related cognitive deficit by N-acetylcysteine in a transgenic mouse model. *Neurobiol Dis, 45*(3), 1111-1120. doi:10.1016/j.nbd.2011.12.031
236. Searcy, J. L., Phelps, J. T., Pancani, T., Kadish, I., Popovic, J., Anderson, K. L., . . . Thibault, O. (2012). Long-term pioglitazone treatment improves learning and attenuates pathological markers in a mouse model of Alzheimer's disease. *J Alzheimers Dis, 30*(4), 943-961. doi:10.3233/JAD-2012-111661
237. Giannopoulos, P. F., Chu, J., Joshi, Y. B., Sperow, M., Li, J. G., Kirby, L. G., & Pratico, D. (2013). 5-lipoxygenase activating protein reduction ameliorates cognitive deficit, synaptic dysfunction, and neuropathology in a mouse model of Alzheimer's disease. *Biol Psychiatry, 74*(5), 348-356. doi:10.1016/j.biopsych.2013.04.009
238. Sancheti, H., Akopian, G., Yin, F., Brinton, R. D., Walsh, J. P., & Cadenas, E. (2013). Age-dependent modulation of synaptic plasticity and insulin mimetic effect of lipoic acid on a mouse model of Alzheimer's disease. *PLoS ONE, 8*(7), e69830. doi:10.1371/journal.pone.0069830
239. Giannopoulos, P. F., Chu, J., Joshi, Y. B., Sperow, M., Li, J. G., Kirby, L. G., & Pratico, D. (2014). Gene knockout of 5-lipoxygenase rescues synaptic dysfunction and improves memory in the triple-transgenic model of Alzheimer's disease. *Mol Psychiatry, 19*(4), 511-518. doi:10.1038/mp.2013.23
240. Grigoryan, G., Biella, G., Albani, D., Forloni, G., & Segal, M. (2014). Stress impairs synaptic plasticity in triple-transgenic Alzheimer's disease mice: rescue by ryanodine. *Neurodegener Dis, 13*(2-3), 135-138. doi:10.1159/000354231
241. Shilling, D., Muller, M., Takano, H., Mak, D. O., Abel, T., Coulter, D. A., & Foskett, J. K. (2014). Suppression of InsP3 receptor-mediated Ca2+ signaling alleviates mutant presenilin-linked familial Alzheimer's disease pathogenesis. *J Neurosci, 34*(20), 6910-6923. doi:10.1523/JNEUROSCI.5441-13.2014
242. Caccamo, A., Branca, C., Talboom, J. S., Shaw, D. M., Turner, D., Ma, L., . . . Oddo, S. (2015). Reducing Ribosomal Protein S6 Kinase 1 Expression Improves Spatial Memory

and Synaptic Plasticity in a Mouse Model of Alzheimer's Disease. *J Neurosci, 35*(41), 14042-14056. doi:10.1523/JNEUROSCI.2781-15.2015
243. Chakroborty, S., Kim, J., Schneider, C., West, A. R., & Stutzmann, G. E. (2015). Nitric oxide signaling is recruited as a compensatory mechanism for sustaining synaptic plasticity in Alzheimer's disease mice. *J Neurosci, 35*(17), 6893-6902. doi:10.1523/JNEUROSCI.4002-14.2015
244. Sykora, P., Misiak, M., Wang, Y., Ghosh, S., Leandro, G. S., Liu, D., . . . Bohr, V. A. (2015). DNA polymerase beta deficiency leads to neurodegeneration and exacerbates Alzheimer disease phenotypes. *Nucleic Acids Res, 43*(2), 943-959. doi:10.1093/nar/gku1356
245. Wang, F., Zhang, Y., Wang, L., Sun, P., Luo, X., Ishigaki, Y., . . . Kato, N. (2015). Improvement of spatial learning by facilitating large-conductance calcium-activated potassium channel with transcranial magnetic stimulation in Alzheimer's disease model mice. *Neuropharmacology, 97*, 210-219. doi:10.1016/j.neuropharm.2015.05.027
246. Wang, L., Kang, H., Li, Y., Shui, Y., Yamamoto, R., Sugai, T., & Kato, N. (2015). Cognitive recovery by chronic activation of the large-conductance calcium-activated potassium channel in a mouse model of Alzheimer's disease. *Neuropharmacology, 92*, 8-15. doi:10.1016/j.neuropharm.2014.12.033
247. Luo, J., Lee, S. H., VandeVrede, L., Qin, Z., Ben Aissa, M., Larson, J., . . . Thatcher, G. R. (2016). A multifunctional therapeutic approach to disease modification in multiple familial mouse models and a novel sporadic model of Alzheimer's disease. *Mol Neurodegener, 11*, 35. doi:10.1186/s13024-016-0103-6
248. Cisse, M., Duplan, E., Lorivel, T., Dunys, J., Bauer, C., Meckler, X., . . . Checler, F. (2017). The transcription factor XBP1s restores hippocampal synaptic plasticity and memory by control of the Kalirin-7 pathway in Alzheimer model. *Mol Psychiatry, 22*(11), 1562-1575. doi:10.1038/mp.2016.152
249. Jin, L., Gao, L. F., Sun, D. S., Wu, H., Wang, Q., Ke, D., . . . Liu, G. P. (2017). Long-term Ameliorative Effects of the Antidepressant Fluoxetine Exposure on Cognitive Deficits in 3 x TgAD Mice. *Mol Neurobiol, 54*(6), 4160-4171. doi:10.1007/s12035-016-9952-9
250. Sun, D. S., Gao, L. F., Jin, L., Wu, H., Wang, Q., Zhou, Y., . . . Liu, G. P. (2017). Fluoxetine administration during adolescence attenuates cognitive and synaptic deficits in adult 3xTgAD mice. *Neuropharmacology, 126*, 200-212. doi:10.1016/j.neuropharm.2017.08.037
251. Wang, Z. H., Liu, P., Liu, X., Manfredsson, F. P., Sandoval, I. M., Yu, S. P., . . . Ye, K. (2017). Delta-Secretase Phosphorylation by SRPK2 Enhances Its Enzymatic Activity, Provoking Pathogenesis in Alzheimer's Disease. *Mol Cell, 67*(5), 812-825 e815. doi:10.1016/j.molcel.2017.07.018
252. Zhao, Y., Sivaji, S., Chiang, M. C., Ali, H., Zukowski, M., Ali, S., . . . Wills, Z. P. (2017). Amyloid Beta Peptides Block New Synapse Assembly by Nogo Receptor-Mediated Inhibition of T-Type Calcium Channels. *Neuron, 96*(2), 355-372 e356. doi:10.1016/j.neuron.2017.09.041
253. Gelman, S., Palma, J., Tombaugh, G., & Ghavami, A. (2018). Differences in Synaptic Dysfunction Between rTg4510 and APP/PS1 Mouse Models of Alzheimer's Disease. *J Alzheimers Dis, 61*(1), 195-208. doi:10.3233/JAD-170457
254. Gong, B., Vitolo, O. V., Trinchese, F., Liu, S., Shelanski, M., & Arancio, O. (2004). Persistent improvement in synaptic and cognitive functions in an Alzheimer mouse

model after rolipram treatment. *J Clin Invest, 114*(11), 1624-1634. doi:10.1172/JCI22831
255. Battaglia, F., Wang, H. Y., Ghilardi, M. F., Gashi, E., Quartarone, A., Friedman, E., & Nixon, R. A. (2007). Cortical plasticity in Alzheimer's disease in humans and rodents. *Biol Psychiatry, 62*(12), 1405-1412. doi:10.1016/j.biopsych.2007.02.027
256. Liu, L., Orozco, I. J., Planel, E., Wen, Y., Bretteville, A., Krishnamurthy, P., . . . Duff, K. (2008). A transgenic rat that develops Alzheimer's disease-like amyloid pathology, deficits in synaptic plasticity and cognitive impairment. *Neurobiol Dis, 31*(1), 46-57. doi:10.1016/j.nbd.2008.03.005
257. Francis, Y. I., Fa, M., Ashraf, H., Zhang, H., Staniszewski, A., Latchman, D. S., & Arancio, O. (2009). Dysregulation of histone acetylation in the APP/PS1 mouse model of Alzheimer's disease. *J Alzheimers Dis, 18*(1), 131-139. doi:10.3233/JAD-2009-1134
258. Calella, A. M., Farinelli, M., Nuvolone, M., Mirante, O., Moos, R., Falsig, J., . . . Aguzzi, A. (2010). Prion protein and Abeta-related synaptic toxicity impairment. *EMBO Mol Med, 2*(8), 306-314. doi:10.1002/emmm.201000082
259. Bachstetter, A. D., Norris, C. M., Sompol, P., Wilcock, D. M., Goulding, D., Neltner, J. H., . . . Van Eldik, L. J. (2012). Early stage drug treatment that normalizes proinflammatory cytokine production attenuates synaptic dysfunction in a mouse model that exhibits age-dependent progression of Alzheimer's disease-related pathology. *J Neurosci, 32*(30), 10201-10210. doi:10.1523/JNEUROSCI.1496-12.2012
260. Fiorito, J., Saeed, F., Zhang, H., Staniszewski, A., Feng, Y., Francis, Y. I., . . . Arancio, O. (2013). Synthesis of quinoline derivatives: discovery of a potent and selective phosphodiesterase 5 inhibitor for the treatment of Alzheimer's disease. *Eur J Med Chem, 60*, 285-294. doi:10.1016/j.ejmech.2012.12.009
261. Rudobeck, E., Bellone, J. A., Szucs, A., Bonnick, K., Mehrotra-Carter, S., Badaut, J., . . . Vlkolinsky, R. (2017). Low-dose proton radiation effects in a transgenic mouse model of Alzheimer's disease - Implications for space travel. *PLoS ONE, 12*(11), e0186168. doi:10.1371/journal.pone.0186168
262. Roder, S., Danober, L., Pozza, M. F., Lingenhoehl, K., Wiederhold, K. H., & Olpe, H. R. (2003). Electrophysiological studies on the hippocampus and prefrontal cortex assessing the effects of amyloidosis in amyloid precursor protein 23 transgenic mice. *Neuroscience, 120*(3), 705-720. doi:10.1016/s0306-4522(03)00381-6
263. Balducci, C., Tonini, R., Zianni, E., Nazzaro, C., Fiordaliso, F., Salio, M., . . . Forloni, G. (2010). Cognitive deficits associated with alteration of synaptic metaplasticity precede plaque deposition in AbetaPP23 transgenic mice. *J Alzheimers Dis, 21*(4), 1367-1381. doi:10.3233/jad-2010-100675
264. Moriguchi, S., Ishizuka, T., Yabuki, Y., Shioda, N., Sasaki, Y., Tagashira, H., . . . Fukunaga, K. (2018). Blockade of the KATP channel Kir6.2 by memantine represents a novel mechanism relevant to Alzheimer's disease therapy. *Mol Psychiatry, 23*(2), 211-221. doi:10.1038/mp.2016.187
265. Larson, J., Lynch, G., Games, D., & Seubert, P. (1999). Alterations in synaptic transmission and long-term potentiation in hippocampal slices from young and aged PDAPP mice. *Brain Research, 840*(1-2), 23-35. doi:10.1016/s0006-8993(99)01698-4
266. Hartman, R. E., Izumi, Y., Bales, K. R., Paul, S. M., Wozniak, D. F., & Holtzman, D. M. (2005). Treatment with an amyloid-beta antibody ameliorates plaque load, learning deficits, and hippocampal long-term potentiation in a mouse model of Alzheimer's disease. *J Neurosci, 25*(26), 6213-6220. doi:10.1523/JNEUROSCI.0664-05.2005

6. Modelling and treatment characteristics of included studies

ID	Author	Year	Sex(es)	Model(s)	Treatment(s)
101	Chapman	1999	NR	Tg2576	
102	Jacobsen	2006	Male	Tg2576	
103	Ma	2007	NR	Tg2576	
104	Jacobsen	2008	Male	Tg2576	PAZ-417
105	Hermann	2009	Female	Tg2576	
106	Kiyota	2009	NR	Tg2576	
107	Gong	2010	NR	Tg2576	Adenoviral Fbx2 vectors
108	Ma	2010	NR	Tg2576	
109	Townsend	2010	Male	Tg2576	
110	Witton	2010	Male	Tg2576	
111	Balducci	2011	Female	Tg2576	CHF5074
112	D'Amelio	2011	Male	Tg2576	
113	Jung	2011	Male	Tg2576	
114	Wen	2011	NR	Tg2576	
115	Wang	2012	Female	Tg2576	
116	Cavallucci	2013	Male	Tg2576	DHPG
117	Gong	2013	NR	Tg2576	Human intravenous immunoglobulin
118	Gong	2013	Mixed group	Tg2576	
119	Ho	2013	NR	Tg2576	
120	An	2014	Male	Tg2576	Neuritin

121	Lawrence	2014	Male	Tg2576	
122	Lee	2014	Male	Tg2576	
123	Jiang	2015	Mixed group	Tg2576	Adapted MWM training
124	Lante	2015	Male	Tg2576	S-3,5-Dihydroxyphenylglycine + 2R-amino-5-phosphonovaleric acid,
125	Murphy	2015	Mixed group	Tg2576	
126	Severini	2015	Male	Tg2576	
127	Fernandez-Fernandez	2016	Female	Tg2576	
128	Huh	2016	Male	Tg2576	NRG1
129	Shah	2016	Female	Tg2576	
130	Nobili	2017	Male	Tg2576	Selegiline (R-(−)-deprenyl hydrochloride
131	Ricciarelli	2017	Male	Tg2576	GEBR-32a
132	Wang	2017	Male	Tg2576	Adeno-associated viruses (AAVs) for PTPN1
133	Arancio	2004	Male	J20	
134	Saura	2005	NR	J20	
135	Palop	2007	Mixed group	J20	
136	Du	2008	NR	J20	
137	Sun	2008	NR, Male	J20	
138	Harris	2010	NR, Male	J20	
139	Cisse	2011	NR	J20	sh-EphB2-306
140	Kiyota	2011	NR	J20	AAV2/ 1-FGF2
141	Sanchez	2012	Mixed group	J20	Levetiracetam
142	Dubal	2015	Mixed group	J20	
143	Fang	2015	Mixed group	J20	

144	Letronne	2016	NR	J20	
145	Qu	2017	NR	J20	
146	Zhang	2017	NR	J20	
147	Kimura	2009	NR	5xFAD	
148	Kimura	2010	NR	5xFAD	
149	Chen	2012	Mixed group	5xFAD	JZL184
150	Crouzin	2013	Male	5xFAD	
151	Seo	2014	Male	5xFAD	
152	Wu	2014	Mixed group	5xFAD	
153	Zhang	2014	Female	5xFAD	LV-MicroRNA-188-3p
154	Zhang	2015	Male	5xFAD	
155	Baranger	2016	Male	5xFAD	
156	Lee	2016	Male	5xFAD	miR-188-5p overexpression (lentivirus)
157	Colie	2017	Male	5xFAD	
158	Duran-Aniotz	2017	NR	5xFAD	
159	Hwang	2017	Mixed group	5xFAD	
160	MacPherson	2017	Female	5xFAD	XPro159
161	Maezawa	2017	NR	5xFAD	TP70
162	Zhang	2017	Male	5xFAD	Compound 11
163	Zhen	2017	Female	5xFAD	Deep brain stimulation
164	Trinchese	2004	Male	APPswe/PSEN1dE9	
165	Trinchese	2008	NR	APPswe/PSEN1dE9	E64
166	Yoshiike	2008	Male	APPswe/PSEN1dE9	

167	Puzzo	2009	Male	APPswe/PSEN1dE9	Sildenafil
168	Gimbel	2010	NR, Mixed group	APPswe/PSEN1dE9	
169	Volianskis	2010	Female	APPswe/PSEN1dE9	
170	Hsiao	2011	Male	APPswe/PSEN1dE9	Calpetin
171	Inestrosa	2011	NR	APPswe/PSEN1dE9	IDN5706
172	Ma	2011	NR	APPswe/PSEN1dE9	
173	Soderman	2011	Female	APPswe/PSEN1dE9	
174	Furman	2012	Male	APPswe/PSEN1dE9	Gfa2-VIVIT
175	Heneka	2013	NR	APPswe/PSEN1dE9	
176	Inestrosa	2013	Male	APPswe/PSEN1dE9	4-phenylbutyrate
177	Kelly	2013	NR	APPswe/PSEN1dE9	
178	Ma	2013	Mixed group	APPswe/PSEN1dE9	
179	Toth	2013	Male	APPswe/PSEN1dE9	
180	Fu	2014	Mixed group	APPswe/PSEN1dE9	KYL peptide
181	Hong	2014	NR	APPswe/PSEN1dE9	
182	Kummer	2014	Male, Mixed group	APPswe/PSEN1dE9	
183	Ma	2014	Mixed group	APPswe/PSEN1dE9	
184	Metais	2014	Mixed group	APPswe/PSEN1dE9	Simvastatin
185	Serrano	2014	Male	APPswe/PSEN1dE9	Andrographolide
186	Vargas	2014	Male	APPswe/PSEN1dE9	Formylated Wnt-5a-derived hexapeptide (FOXY-5)
187	Vegh	2014	Male	APPswe/PSEN1dE9	Chondroitinase ABC
188	Zhang	2014	NR	APPswe/PSEN1dE9	Neural stem cells from WT mice
189	Zhang	2014	NR	APPswe/PSEN1dE9	Lamotrigine

190	Zhao	2014	Male	APPswe/PSEN1dE9	Tri-lithium pyrroloquinoline quinone (Li3PQQ)
191	Cisternas	2015	Male	APPswe/PSEN1dE9	Potassium chloride
192	Hu	2015	Male	APPswe/PSEN1dE9	
193	Kim	2015	Male	APPswe/PSEN1dE9	4-(2-hydroxyethyl)-1-piperazinepropanesulphonic acid
194	Megill	2015	Male, Mixed group	APPswe/PSEN1dE9	
195	Woo	2015	Male	APPswe/PSEN1dE9	
196	Woo	2015	Mixed group	APPswe/PSEN1dE9	
197	Fu	2016	Mixed group	APPswe/PSEN1dE9	IL-33
198	Kajiwara	2016	Male	APPswe/PSEN1dE9	
199	Li	2016	Male	APPswe/PSEN1dE9	Ginsenoside Rg1
200	Montgomery	2016	Female	APPswe/PSEN1dE9	
201	Oyelami	2016	NR	APPswe/PSEN1dE9	
202	Shen	2016	NR, Male	APPswe/PSEN1dE9	D-Tyr MTII (D-Tyr),
203	Yang	2016	Mixed group	APPswe/PSEN1dE9	
204	Yang	2016	Male	APPswe/PSEN1dE9	Sodium Hydrosulfide
205	Alves	2017	Male	APPswe/PSEN1dE9	AAV8-CAG-IL2
206	Gomez-Gonzalo	2017	Mixed group	APPswe/PSEN1dE9	
207	Jin	2017	Female	APPswe/PSEN1dE9	Peritoneal dialysis surgery
208	Li	2017	Male	APPswe/PSEN1dE9	
209	Maezawa	2017	NR	APPswe/PSEN1dE9	PAP1 (5-(4-phenoxybutoxy)psoralen)
210	Tao	2017	Male	APPswe/PSEN1dE9	Lenti-Flag-HDAC1WT-SUMO1
211	Yu	2017	Male	APPswe/PSEN1dE9	PSD-93 overexpresssing lentivirus

212	Martinez Hernandez	2018	Male	APPswe/PSEN1dE9	Anle138b
213	Wang	2018	Male	APPswe/PSEN1dE9	Mesenchymal stem cells-derived extracellular vesicles
214	Zhu	2018	Male	APPswe/PSEN1dE9	Xanthoceraside
215	Gureviciene	2004	Male	APPSwe/PSEN1(A246E)	
216	Fitzjohn	2010	Mixed group	APPSwe/PSEN1(A246E)	
217	Ricoy	2011	NR	APPSwe/PSEN1(A246E)	
218	Maingret	2017	Male	APPSwe/PSEN1(A246E)	
219	Jolas	2002	NR	TgCRND8	
220	Arrieta-Cruz	2010	Male	TgCRND8	
221	Arrieta-Cruz	2010	Male	TgCRND8	
222	Ye	2010	NR	TgCRND8	
223	Sclip	2011	NR	TgCRND8	D-JNKI1
224	Sclip	2014	NR	TgCRND8	
225	Luccarini	2015	NR	TgCRND8	
226	Tozzi	2015	NR	TgCRND8	
227	Cavanagh	2016	Male	TgCRND8	XPRO1595
228	Hinrich	2016	Female, Male, Mixed group	TgCRND8	ASO-21
229	Kimura	2017	Mixed group	TgCRND8	
230	Knock	2018	Mixed group	TgCRND8	
231	Oddo	2003	NR	3xTg-AD	
232	Chakroborty	2009	Mixed group	3xTg-AD	
233	Chakroborty	2012	Mixed group	3xTg-AD	
234	Chakroborty	2012	Mixed group	3xTg-AD	Dantrolene

235	Hsiao	2012	Male	3xTg-AD	N-acetylcysteine
236	Searcy	2012	Female	3xTg-AD	Pioglitazone (PIO)
237	Giannopoulos	2013	Mixed group	3xTg-AD	MK-591
238	Sancheti	2013	NR	3xTg-AD	R sodium lipoic acid
239	Giannopoulos	2014	Mixed group	3xTg-AD	Zileuton
240	Grigoryan	2014	Male	3xTg-AD	Stress
241	Shilling	2014	Male	3xTg-AD	
242	Caccamo	2015	Female	3xTg-AD	
243	Chakroborty	2015	Mixed group	3xTg-AD	
244	Sykora	2015	Female	3xTg-AD	
245	Wang	2015	Male , Female	3xTg-AD	TMS stimulation
246	Wang	2015	Male	3xTg-AD	Isopimaric acid
247	Luo	2016	Male	3xTg-AD	NMZ
248	Cisse	2017	Female	3xTg-AD	Lentiviral vectors with scrambled shRNA + XBP1s
249	Jin	2017	Mixed group	3xTg-AD	Fluoxetine
250	Sun	2017	Mixed group	3xTg-AD	Fluoxetine
251	Wang	2017	Mixed group	3xTg-AD	AEP-WT
252	Zhao	2017	Mixed group	3xTg-AD	
253	Gelman	2018	Male	PS/APP	
254	Gong	2004	Mixed group	APPPS1	Rolipram
255	Battaglia	2007	Mixed group	APPPS1	
256	Liu	2008	NR, Mixed group	APPPS1	
257	Francis	2009	Mixed group	APPPS1	Trichostatin A

258	Calella	2010	Male	APPPS1	
259	Bachstetter	2012	Mixed group	APPPS1	MW-151
260	Fiorito	2013	Mixed group	APPPS1	Compound 7a
261	Rudobeck	2017	Male	APPPS1	Radiation
262	Roder	2003	Male	APP23	
263	Balducci	2010	Male	APP23	
264	Moriguchi	2016	Male	APP23	Memantine
265	Larson	1999	NR	PDAPP	
266	Hartman	2005	Mixed group	PDAPP	Anti-AB antibody 10D5

NR: Not reported.

7. Slicing and recording protocols by publication

ID	Model	Time of day sacrificed	Anesthetised prior to sacrifice	Time slices left to recover (mins)	Temp slices left to recover (*C)	Kyn acid	Mg^{2+} slicing solution	Ca^{2+} slicing solution	Pathway(s) recorded	Type of recording chamber	Ca^{2+} recording solution	Mg^{2+} recording solution
101	Tg2576	NR	NR	60	RT	NR	1.3	2.5	Schaffer collaterals, Perforant path	Submersion	NR	NR
102	Tg2576	NR	NR	NR	NR	NR	NR	NR	Perforant path	NR	NR	NR
103	Tg2576	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	2.5	1.3
104	Tg2576	NR	NR	NR	NR	NR	NR	NR	Perforant path	NR	NR	NR
105	Tg2576	NR	NR	60	33	NR	2	2.5	Schaffer collaterals	Interface	NR	NR
106	Tg2576	NR	Yes - Isoflurane	60	RT	NR	NR	NR	Schaffer collaterals	Submersion	2	2
107	Tg2576	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
108	Tg2576	NR	Yes - Isoflurane	120	RT	NR	1.3	2.5	Schaffer collaterals	Submersion	NR	NR
109	Tg2576	NR	Yes - Isoflurane	120	RT	Yes	3	0	Schaffer collaterals	NR	0	2
110	Tg2576	NR	No - cervical dislocation	60	NR	NR	NR	NR	Mossy fiber pathway	Interface	2	1
111	Tg2576	NR	Yes - Halothane	NR	NR	NR	NR	NR	Schaffer collaterals	Submersion	2.5	1.3
112	Tg2576	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	2	2
113	Tg2576	NR	NR	NR	RT	NR	1.3	2.5	Mossy fibre pathway, Schaffer collaterals	NR	NR	NR
114	Tg2576	NR	NR	330	34; RT	NR	10	0.5	Schaffer collaterals	NR	2	2
115	Tg2576	NR	NR	270	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
116	Tg2576	NR	Yes - 2-Bromo-2-Chloro-1,1,1-trifluoroethane	60	RT	NR	1	2	Schaffer collaterals	NR	NR	NR
117	Tg2576	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
118	Tg2576	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2

119	Tg2576	NR	NR	120	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
120	Tg2576	NR	NR	60	37	NR	1.3	0	Schaffer collaterals	Submersion	NR	NR
121	Tg2576	NR	Yes - Tribromoethanol	60	RT; 32	NR	1.5	2	Schaffer collaterals	NR	NR	NR
122	Tg2576	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	NR	2	2
123	Tg2576	NR	Yes - Chloral hydrate	NR	NR	NR	1	2	Schaffer collaterals	NR	2	1
124	Tg2576	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
125	Tg2576	NR	Yes - Isoflurane	60	RT	NR	7	1	Schaffer collaterals	NR	NR	NR
126	Tg2576	NR	NR	60	NR	NR	NR	NR	Schaffer collaterals	Submersion	2.4	1.2
127	Tg2576	NR	Yes - Isoflurane	60	RT	NR	8	0	Schaffer collaterals	NR	2.25	1.2
128	Tg2576	NR	NR	60	NR	NR	1	2	Schaffer collaterals	NR	NR	NR
129	Tg2576	NR	NR	150	32	NR	2	2	Schaffer collaterals	Submersion	NR	NR
130	Tg2576	NR	Yes	90	32; RT	NR	1	2	Schaffer collaterals	MEA	NR	NR
131	Tg2576	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
132	Tg2576	NR	NR	90	32	NR	1.2	2.5	Schaffer collaterals	MEA	NR	NR
133	J20	NR	NR	90	NR	NR	NR	NR	Schaffer collaterals	Interface	2	2
134	J20	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
135	J20	NR	NR	60	35	NR	7	0.5	Schaffer collaterals, Perforant path	NR	2	1
136	J20	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	Interface	2	2
137	J20	NR	NR	45	RT	NR	10	0.5	Schaffer collaterals, Perforant path	Submersion	NR	NR
138	J20	NR	Yes - Avertin	60	30; RT	NR	10	0.5	Schaffer collaterals, Perforant path	Submersion	2	1
139	J20	NR	Yes - Avertin	60	30; RT	NR	10	0.5	Schaffer collaterals, Perforant path	Submersion	2	1
140	J20	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
141	J20	NR	NR	90	32; RT	NR	10	0	Perforant path	Submersion	2	1
142	J20	NR	Yes - Isoflurane	60	RT	NR	10	0	Perforant path	Submersion	NR	NR

143	J20	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	Interface	2	2
144	J20	NR	NR	80	34; RT	NR	NR	NR	Schaffer collaterals	Submersion	2.5	1.3
145	J20	NR	NR	90	29	NR	22	2	Schaffer collaterals	Interface	NR	NR
146	J20	NR	NR	70	34; RT	NR	10	0.5	Perforant path	Submersion	2	1
147	5xFAD	NR	Yes - Isoflurane	60	RT	NR	NR	NR	Schaffer collaterals	Submersion	2.4	2
148	5xFAD	NR	Yes - Isoflurane	60	23	NR	NR	NR	Schaffer collaterals	Submersion	2.4	2
149	5xFAD	NR	NR	120	36; 23	NR	1	2	Schaffer collaterals, Perforant path	NR	NR	NR
150	5xFAD	NR	NR	60	RT	NR	2	1	Schaffer collaterals	MEA	2	2
151	5xFAD	NR	NR	60	29	NR	10	0.5	Schaffer collaterals	NR	2.5	1.5
152	5xFAD	NR	Yes - Avertin	60	32; RT	NR	10	0.5	Perforant path	Submersion	2	1.3
153	5xFAD	NR	NR	135	36; 23	NR	NR	NR	Schaffer collaterals	NR	NR	NR
154	5xFAD	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
155	5xFAD	NR	NR	60	RT	NR	2	1	Schaffer collaterals	MEA	2	2
156	5xFAD	NR	NR	60	36	NR	3.5	0.5	Schaffer collaterals	NR	2.5	1.3
157	5xFAD	NR	NR	60	32	NR	5	0	Schaffer collaterals	NR	2.5	1.2
158	5xFAD	NR	Yes - Isoflurane	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
159	5xFAD	NR	NR	60	NR	NR	NR	NR	Schaffer collaterals	NR	2.5	1.3
160	5xFAD	NR	Yes- CO2	90	32	NR	2	0	Schaffer collaterals	NR	2	2
161	5xFAD	NR	Yes - Isoflurane	60	RT	NR	1.3	2.4	Schaffer collaterals	Submersion	NR	NR
162	5xFAD	NR	Yes - Isoflurane	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
163	5xFAD	NR	Yes - Isoflurane	40	35	NR	2.4	2.5	Perforant path	MEA	NR	NR
164	APPswe/PSEN1dE9	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	NR	NR	NR
165	APPswe/PSEN1dE9	NR	NR	90	NR	NR	NR	NR	Schaffer collaterals	Interface	NR	NR
166	APPswe/PSEN1dE9	NR	NR	60	RT	NR	2	2	Perforant path	NR	NR	NR
167	APPswe/PSEN1dE9	NR	No - cervical dislocation	NR	NR	NR	NR	NR	Schaffer collaterals	NR	2	2
168	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
169	APPswe/PSEN1dE9	Morning	Yes - Halothane	120	RT	NR	NR	NR	Schaffer collaterals	NR	2	2
170	APPswe/PSEN1dE9	NR	NR	60	NR	NR	1.2	2.5	Schaffer collaterals	NR	NR	NR

171	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
172	APPswe/PSEN1dE9	NR	NR	120	NR	NR	NR	NR	Schaffer collaterals	NR	2	1
173	APPswe/PSEN1dE9	Daytime	NR	60	RT	NR	NR	NR	Schaffer collaterals	Submersion	2	2
174	APPswe/PSEN1dE9	NR	Yes- CO2	90	32	NR	2	0	Schaffer collaterals	Interface	0	2
175	APPswe/PSEN1dE9	NR	Yes	90	RT	NR	2	1.25	Schaffer collaterals	Submersion	NR	NR
176	APPswe/PSEN1dE9	NR	NR	60	RT	NR	1.3	2.5	Schaffer collaterals	NR	NR	NR
177	APPswe/PSEN1dE9	NR	NR	NR	22	NR	NR	NR	Schaffer collaterals	Submersion	2	1.5
178	APPswe/PSEN1dE9	NR	NR	120	NR	NR	NR	NR	Schaffer collaterals	NR	2	1
179	APPswe/PSEN1dE9	NR	NR	60	ambient temperature	NR	NR	NR	Schaffer collaterals	MEA	3.5	2
180	APPswe/PSEN1dE9	NR	NR	120	32	NR	NR	NR	Schaffer collaterals	MEA	NR	NR
181	APPswe/PSEN1dE9	NR	Yes - Ether	120	32	NR	11.3	1.5	Schaffer collaterals	Interface	2.5	1.5
182	APPswe/PSEN1dE9	NR	Yes - Isoflurane	60	29	NR	NR	NR	Schaffer collaterals	NR	NR	NR
183	APPswe/PSEN1dE9	NR	NR	120	RT	NR	NR	NR	Schaffer collaterals	NR	NR	NR
184	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
185	APPswe/PSEN1dE9	NR	NR	60	RT	NR	1.3	2.5	Schaffer collaterals	NR	NR	NR
186	APPswe/PSEN1dE9	NR	NR	60	23	NR	NR	NR	Schaffer collaterals	NR	NR	NR
187	APPswe/PSEN1dE9	NR	NR	60	NR	NR	7	0.5	Schaffer collaterals	MEA	2.5	1.3
188	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
189	APPswe/PSEN1dE9	NR	NR	60	NR	NR	NR	NR	Schaffer collaterals	Submersion	NR	NR
190	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
191	APPswe/PSEN1dE9	NR	NR	60	NR	NR	NR	NR	NR	NR	NR	NR
192	APPswe/PSEN1dE9	NR	Yes - Ether	60	RT	NR	1.5	2.5	Schaffer collaterals	NR	NR	NR
193	APPswe/PSEN1dE9	NR	Yes - Isoflurane	90	35; 24	Yes	0	0	Schaffer collaterals	NR	2.5	1.3
194	APPswe/PSEN1dE9	NR	Yes - Isoflurane	NR	RT	NR	3	1	Schaffer collaterals	Submersion	2.5	1.5
195	APPswe/PSEN1dE9	NR	NR	100	RT; 30	NR	7	0.5	Schaffer collaterals	NR	2	1.2
196	APPswe/PSEN1dE9	NR	NR	100	RT; 30	NR	7	0.5	Schaffer collaterals	NR	2	1.2
197	APPswe/PSEN1dE9	Daytime	NR	120	32	NR	NR	NR	Schaffer collaterals	MEA	NR	NR
198	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	2.5	1.3

199	APPswe/PSEN1dE9	NR	NR	120	32	NR	5	3.4	Schaffer collaterals	Interface	2.5	1.5
200	APPswe/PSEN1dE9	NR	Yes - Halothane	NR	NR	NR	NR	NR	Schaffer collaterals	Interface	2	2
201	APPswe/PSEN1dE9	NR	NR	60	RT	NR	2.5	2	Schaffer collaterals	MEA	2	1
202	APPswe/PSEN1dE9	NR	NR	120	32	NR	NR	NR	Schaffer collaterals	MEA	NR	NR
203	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
204	APPswe/PSEN1dE9	NR	NR	90	30	NR	NR	NR	Schaffer collaterals	NR	NR	NR
205	APPswe/PSEN1dE9	NR	Yes- CO2	120	32	NR	2	2	Schaffer collaterals	NR	NR	NR
206	APPswe/PSEN1dE9	NR	Yes	60	RT	NR	NR	NR	Schaffer collaterals	NR	NR	NR
207	APPswe/PSEN1dE9	NR	Yes - Chloral hydrate	30	30	NR	NR	NR	Schaffer collaterals	NR	NR	NR
208	APPswe/PSEN1dE9	NR	Yes- CO2	150	32	NR	NR	NR	Schaffer collaterals	Interface	2	2
209	APPswe/PSEN1dE9	NR	Yes	40	35	NR	6	0.2	Schaffer collaterals	NR	2	1
210	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
211	APPswe/PSEN1dE9	NR	Yes - Servoflurane	120	30; then RT	NR	7	0.5	Schaffer collaterals	NR	2.5	1.3
212	APPswe/PSEN1dE9	NR	Yes- CO2	180	32	NR	NR	NR	Schaffer collaterals	Interface	2	2
213	APPswe/PSEN1dE9	NR	Yes - Ether	60	32	NR	10	1	Schaffer collaterals	NR	2.5	1.3
214	APPswe/PSEN1dE9	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
215	APPSwe/PSEN1(A246E)	NR	Yes - Halothane	60	35	NR	NR	NR	Schaffer collaterals	Interface	3.4	2.5
216	APPSwe/PSEN1(A246E)	NR	NR	60	33	Yes	1.2	1.4	Schaffer collaterals	Submersion	NR	NR
217	APPSwe/PSEN1(A246E)	NR	Yes - Isoflurane	NR	NR	NR	7	0.5	Schaffer collaterals	NR	4	4
218	APPSwe/PSEN1(A246E)	NR	Yes - Ketamine	30	33; Room termperature	NR	8	0.4	Perforant path	Submersion	2	1
219	TgCRND8	NR	NR	120	RT	NR	2.4	2.5	Schaffer collaterals	Submersion	NR	NR
220	TgCRND8	NR	NR	120	RT	NR	NR	NR	Schaffer collaterals	NR	1	4

221	TgCRND8	NR	NR	120	RT	NR	NR	NR	Schaffer collaterals	NR	1	4
222	TgCRND8	NR	Yes - Ketamine	60	RT	NR	4	1	Schaffer collaterals	NR	2	2
223	TgCRND8	NR	NR	120	30; RT	NR	1.2	2.4	Schaffer collaterals	Submersion	NR	NR
224	TgCRND8	NR	NR	120	30; RT	NR	1.2	2.4	Perforant path	Submersion	NR	NR
225	TgCRND8	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	2.4	2
226	TgCRND8	NR	NR	120	30; RT	NR	1.2	2.4	Schaffer collaterals	Submersion	NR	NR
227	TgCRND8	NR	Yes - Isoflurane	NR	32	NR	1	2	Schaffer collaterals	NR	NR	NR
228	TgCRND8	NR	Yes - Halothane	NR	32	NR	4	0.5	Schaffer collaterals	Submersion	2	1.2
229	TgCRND8	NR	NR	60	RT	NR	NR	NR	Schaffer collaterals	Submersion	2.4	2
230	TgCRND8	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
231	3xTg-AD	NR	Yes - Halothane	60	RT	NR	1.2	2	Schaffer collaterals	Interface	NR	NR
232	3xTg-AD	NR	Yes - Halothane	NR	27	NR	1.2	2	Schaffer collaterals	Interface	NR	NR
233	3xTg-AD	NR	NR	NR	27	NR	1.2	2	Schaffer collaterals	Interface	NR	NR
234	3xTg-AD	NR	Yes - Halothane	NR	27	NR	1.2	2	Schaffer collaterals	Interface	2	1.2
235	3xTg-AD	NR	NR	60	NR	NR	NR	NR	Schaffer collaterals	NR	2.5	1.2
236	3xTg-AD	NR	Yes- CO2	120	32	NR	8	0.1	Schaffer collaterals	Interface	2.5	1.3
237	3xTg-AD	NR	NR	60	RT	NR	2	NR	Schaffer collaterals	NR	NR	NR
238	3xTg-AD	NR	Yes - Isoflurane	60	RT	NR	4	0	Schaffer collaterals	Interface	1.3	2.4
239	3xTg-AD	NR	NR	60	RT	NR	2	2.5	Schaffer collaterals	NR	2.5	2
240	3xTg-AD	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
241	3xTg-AD	NR	No - cervical dislocation	120	28	NR	NR	NR	Schaffer collaterals	Interface	2.5	1.3
242	3xTg-AD	NR	Yes - Isoflurane	60	RT	NR	1.3	2.5	Schaffer collaterals	Interface	NR	NR
243	3xTg-AD	NR	Yes	60	30	NR	4	0.5	Schaffer collaterals	Interface	NR	NR
244	3xTg-AD	NR	Yes - Isoflurane	60	NR	NR	NR	NR	Schaffer collaterals	NR	2.5	1.3
245	3xTg-AD	NR	Yes - Ether	NR	25	NR	2	2.5	Schaffer collaterals	NR	2.5	2
246	3xTg-AD	NR	Yes - Ether	NR	NR	NR	2	2.5	Schaffer collaterals	NR	NR	NR
247	3xTg-AD	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
248	3xTg-AD	NR	Yes - Isoflurane	30	29	NR	10	0.5	Schaffer collaterals	Submersion	10	0.5
249	3xTg-AD	NR	NR	30	NR	NR	2	1.25	Perforant path	MEA	NR	NR

250	3xTg-AD	NR	Yes	30	NR	NR	NR	NR	Perforant path	Submersion	1.25	2
251	3xTg-AD	NR	Yes - Isoflurane	75	23.5	NR	6	1	Schaffer collaterals	NR	NR	NR
252	3xTg-AD	NR	Yes - Isoflurane	NR	RT	NR	1.2	2	Schaffer collaterals	Interface	NR	NR
253	PS/APP	NR	Yes - Isoflurane	90	32	NR	1.2	2.5	Schaffer collaterals	Interface	NR	NR
254	APPPS1	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
255	APPPS1	NR	NR	60	27.5	NR	NR	NR	Perforant path	Interface	2.5	1.3
256	APPPS1	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	NR	NR
257	APPPS1	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
258	APPPS1	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	NR	NR	NR
259	APPPS1	NR	Yes - CO2	105	32	NR	2	0	Schaffer collaterals	NR	2	2
260	APPPS1	NR	NR	90	29	NR	NR	NR	Schaffer collaterals	Interface	2	2
261	APPPS1	NR	NR	80	33	NR	NR	NR	Schaffer collaterals	Submersion	NR	NR
262	APP23	NR	Yes - Isoflurane	60	RT	NR	2	2.5	Schaffer collaterals	Interface	2.5	2
263	APP23	Daytime	Yes - Isoflurane	NR	NR	NR	7	0	Schaffer collaterals	Submersion	2.4	1.3
264	APP23	NR	Yes	120	28	NR	NR	NR	Schaffer collaterals	Interface	NR	NR
265	PDAPP	NR	NR	NR	NR	NR	NR	NR	Schaffer collaterals	Interface	1.2	3.4
266	PDAPP	NR	Yes - Halothane	60	30	NR	2	2	Schaffer collaterals	Submersion	2.5	1.3

NR: not reported.

*8. **MWM protocols by publication***

ID	Visible platform trial(s)	Method of measurement	Acquisition phase time (seconds)	Probe phase time (seconds)	Habituation time	Water temperature °C	Number of training days	Number of trials per day	Total trials
103	Yes	Automatic tracking software	NR	60	NR	NR	9	4	36
114	No	Automatic tracking software	120	60	60 seconds	22	5	4	20
123	No	Automatic tracking software	60	NR	NR	NR	6	4	24
130	Yes	Automatic tracking software	60	30	NR	23	4, 1 (2 protocols)	5, 8	20, 8
132	No	Automatic tracking software	NR	NR	NR	NR	7	NR	
137	No	Automatic tracking software	60	60	6 minutes trial	20	5	6	30
138	No	Automatic tracking software	60	60	360 seconds	21	5	4	20
138	No	Automatic tracking software	60	60	360 seconds	21	5	6	30
143	No	Automatic tracking software	60	60	NR	23	6	4	24
149	Yes	Automatic tracking software	60	60	NR	NR	7	4	28
153	Yes	Automatic tracking software	60	60	NR	NR	3	4	12
153	Yes	Automatic tracking software	60	60	NR	NR	7	4	28
154	No	Automatic tracking software	60	60	NR	23.5	5	4	20
162	No	Automatic tracking software	60	60	NR	23.5	5	4	20
163	Yes	Automatic tracking software	60	60	NR	NR	5	6	30
166	No	Automatic tracking software	60	60	NR	24	9	3	27
167	Yes	Automatic tracking software	60	60	NR	NR	3	2	6
168	No	Automatic tracking software	60	60	NR	NR	3	8	24
168	No	Automatic tracking software	60	60	NR	NR	1	5	5
175	No	Automatic tracking software	40	30	NR	24	8	4	32
176	No	Automatic tracking software	60	60	NR	20	8	3	24
178	Yes	Automatic tracking software	60	NR	NR	NR	5	4	20

179	No	Automatic tracking software	90	NR	NR	22.5	5	2	10
181	No	Automatic tracking software	60	60	NR	NR	8	4	32
182	No	Automatic tracking software	40	NR	NR	24	8	4	32
185	No	Automatic tracking software	60	NA	NR	20	10	3	30
187	No	Automatic tracking software	60	60	NR	25	5	4	20
188	No	Automatic tracking software	60	NR	NR	22	6	NR	
190	No	Automatic tracking software	60	NR	NR	NR	5	4	20
191	No	Automatic tracking software	60	60	NR	20	8	3	24
193	No	Automatic tracking software	60	60	NR	25	6	5	30
211	No	Automatic tracking software	60	60	NR	NR	5	NR	
212	No	Automatic tracking software	NR	NR	NR	NR	8	4	32
213	No	Automatic tracking software	60	60	NR	NR	5	8	40
228	No	Automatic tracking software	60	60	NR	23	4	8	32
242	No	Automatic tracking software	NA	60	NR	25	5	4	20
244	Yes	Automatic tracking software	60	60	NR	25	7	4	28
245	No	Automatic tracking software	60	60	NR	25	5	4	20
246	No	Automatic tracking software	60	60	NR	25	5	4	20
248	Yes	Automatic tracking software	60	60	6 minutes	21	5	4	20
249	No	Automatic tracking software	60	NR	NR	24	6	4	24
250	No	Automatic tracking software	60	NR	NR	24	6	4	24
254	No	Automatic tracking software	NR	NR	NR	NR	3	6	18
256	No	Automatic tracking software	60	60	1 min/day for 3 days	22	5	4	20
266	Yes	Automatic tracking software	60	30 (standard); 60 (modified)	NR	NR	25	4	100

NR: Not reported.

*9. **RoB and study quality reporting by publication***

ID	Author	Year	Exclusion criteria	Blinding	Welfare committee approval	Sample size calculation	Conflict of interest statement	Randomisation	Allocation concealment
101	Chapman	1999	NR	Reported	NR	NR	NR	NA	NA
102	Jacobsen	2006	NR	NR	NR	NR	Reported	NA	NA
103	Ma	2007	Reported	Reported	NR	NR	NR	NA	NA
104	Jacobsen	2008	NR	NR	NR	NR	Reported	NR	NR
105	Hermann	2009	NR	Reported	Reported	NR	NR	NA	NA
106	Kiyota	2009	NR	NR	Reported	NR	NR	NA	NA
107	Gong	2010	NR	NR	NR	NR	Reported	NR	NR
108	Ma	2010	NR	NR	Reported	NR	Reported	NA	NA
109	Townsend	2010	Reported	NR	Reported	NR	NR	NA	NA
110	Witton	2010	NR	Reported	NR	NR	Reported	NA	NA
111	Balducci	2011	NR	Reported	NR	NR	Reported	NR	Reported
112	D'Amelio	2011	NR	NR	Reported	NR	Reported	NA	NA
113	Jung	2011	NR	NR	NR	NR	NR	NA	NA
114	Wen	2011	Reported	Reported	Reported	NR	Reported	NA	NA
115	Wang	2012	NR	NR	Reported	NR	NR	NA	NA
116	Cavallucci	2013	NR	NR	Reported	NR	Reported	NR	NR
117	Gong	2013	NR	NR	Reported	NR	NR	NR	NR
118	Gong	2013	NR	NR	Reported	NR	Reported	NA	NA
119	Ho	2013	NR	NR	Reported	NR	NR	NA	NA
120	An	2014	NR	NR	Reported	NR	Reported	NR	NR
121	Lawrence	2014	NR	NR	Reported	NR	Reported	NA	NA

122	Lee	2014	NR	NR	Reported	NR	Reported	NA	NA
123	Jiang	2015	NR	NR	Reported	NR	Reported	NR	NR
124	Lante	2015	NR	NR	NR	NR	Reported	NR	NR
125	Murphy	2015	Reported	Reported	Reported	NR	Reported	NA	NA
126	Severini	2015	NR	NR	Reported	NR	Reported	NA	NA
127	Fernandez-Fernandez	2016	NR	NR	Reported	NR	Reported	NA	NA
128	Huh	2016	NR	NR	Reported	NR	NR	NR	NR
129	Shah	2016	NR	NR	Reported	NR	NR	NA	NA
130	Nobili	2017	NR	NR	NR	NR	NR	Reported	NR
131	Ricciarelli	2017	NR	NR	Reported	NR	Reported	NR	NR
132	Wang	2017	NR	Reported	Reported	NR	Reported	NR	NR
133	Arancio	2004	NR	NR	NR	NR	NR	NA	NA
134	Saura	2005	NR	Reported	NR	NR	NR	NA	NA
135	Palop	2007	NR	NR	Reported	NR	NR	NA	NA
136	Du	2008	NR	NR	Reported	NR	NR	NA	NA
137	Sun	2008	NR	Reported	NR	NR	NR	NA	NA
138	Harris	2010	NR	NR	Reported	NR	NR	NA	NA
139	Cisse	2011	NR	Reported	NR	NR	Reported	NR	Reported
140	Kiyota	2011	NR	NR	NR	NR	Reported	NR	NR
141	Sanchez	2012	NR	Reported	Reported	NR	NR	NR	NR
142	Dubal	2015	NR	Reported	Reported	NR	NR	NA	NA
143	Fang	2015	NR	NR	Reported	NR	Reported	NA	NA
144	Letronne	2016	NR	Reported	Reported	NR	Reported	NA	NA
145	Qu	2017	NR	NR	Reported	NR	Reported	NA	NA
146	Zhang	2017	NR	NR	NR	NR	Reported	NA	NA
147	Kimura	2009	NR	Reported	Reported	NR	NR	NA	NA
148	Kimura	2010	NR	Reported	Reported	NR	NR	NA	NA
149	Chen	2012	NR	NR	Reported	NR	NR	NR	NR

150	Crouzin	2013	Reported	NR	Reported	NR	Reported	NA	NA
151	Seo	2014	NR	NR	Reported	NR	NR	NA	NA
152	Wu	2014	NR	NR	Reported	NR	NR	NA	NA
153	Zhang	2014	NR	NR	Reported	NR	Reported	NR	NR
154	Zhang	2015	NR	Reported	Reported	Reported	Reported	NA	NA
155	Baranger	2016	Reported	NR	Reported	NR	Reported	NA	NA
156	Lee	2016	NR	Reported	Reported	NR	NR	NR	NR
157	Colie	2017	NR	NR	Reported	NR	NR	NA	NA
158	Duran-Aniotz	2017	NR	NR	Reported	NR	Reported	NA	NA
159	Hwang	2017	NR	Reported	Reported	NR	Reported	NA	NA
160	MacPherson	2017	NR	Reported	Reported	NR	Reported	Reported	NR
161	Maezawa	2017	NR	NR	Reported	NR	Reported	NR	NR
162	Zhang	2017	NR	Reported	Reported	Reported	Reported	Reported	Reported
163	Zhen	2017	NR	Reported	Reported	NR	Reported	NR	NR
164	Trinchese	2004	NR	NR	NR	NR	NR	NA	NA
165	Trinchese	2008	NR	Reported	Reported	Reported	Reported	NR	NR
166	Yoshiike	2008	NR	NR	Reported	NR	Reported	NA	NA
167	Puzzo	2009	NR	Reported	Reported	NR	Reported	NR	Reported
168	Gimbel	2010	NR	Reported	NR	NR	NR	NA	NA
169	Volianskis	2010	Reported	Reported	NR	NR	Reported	NA	NA
170	Hsiao	2011	NR	NR	Reported	NR	Reported	NR	NR
171	Inestrosa	2011	NR	NR	NR	NR	Reported	NR	NR
172	Ma	2011	NR	NR	NR	NR	Reported	NA	NA
173	Soderman	2011	NR	NR	NR	NR	Reported	NA	NA
174	Furman	2012	NR	Reported	NR	NR	Reported	NR	NR
175	Heneka	2013	Reported	Reported	Reported	NR	Reported	NA	NA
176	Inestrosa	2013	NR	NR	NR	NR	NR	NR	NR
177	Kelly	2013	NR	NR	Reported	NR	NR	NA	NA
178	Ma	2013	NR	NR	NR	NR	NR	NA	NA

179	Toth	2013	Reported	NR	Reported	NR	Reported	NA	NA
180	Fu	2014	NR	Reported	Reported	NR	Reported	NR	NR
181	Hong	2014	NR	NR	Reported	NR	Reported	NA	NA
182	Kummer	2014	NR	Reported	Reported	NR	Reported	NA	NA
183	Ma	2014	NR	NR	NR	NR	Reported	NA	NA
184	Metais	2014	NR	NR	Reported	NR	Reported	NR	NR
185	Serrano	2014	NR	NR	Reported	NR	Reported	NR	NR
186	Vargas	2014	NR	NR	NR	NR	Reported	NR	NR
187	Vegh	2014	NR	NR	Reported	NR	Reported	NR	NR
188	Zhang	2014	NR	NR	Reported	NR	NR	NR	NR
189	Zhang	2014	NR	NR	Reported	NR	Reported	Reported	NR
190	Zhao	2014	NR	NR	NR	NR	Reported	Reported	NR
191	Cisternas	2015	NR	NR	NR	NR	Reported	NR	NR
192	Hu	2015	NR	NR	Reported	NR	Reported	NA	NA
193	Kim	2015	NR	NR	NR	NR	NR	NR	NR
194	Megill	2015	NR	NR	Reported	NR	Reported	NA	NA
195	Woo	2015	NR	NR	NR	NR	Reported	NA	NA
196	Woo	2015	NR	NR	NR	NR	Reported	NA	NA
197	Fu	2016	NR	Reported	Reported	NR	Reported	Reported	NR
198	Kajiwara	2016	NR	Reported	Reported	NR	Reported	NA	NA
199	Li	2016	NR	NR	Reported	NR	NR	NR	NR
200	Montgomery	2016	NR	NR	Reported	NR	NR	NA	NA
201	Oyelami	2016	NR	NR	Reported	NR	NR	NA	NA
202	Shen	2016	NR	Reported	Reported	NR	NR	NR	Reported
203	Yang	2016	NR	NR	NR	NR	Reported	NA	NA
204	Yang	2016	NR	NR	Reported	NR	Reported	NR	NR
205	Alves	2017	NR	Reported	Reported	NR	Reported	NR	NR
206	Gomez-Gonzalo	2017	NR	NR	Reported	NR	Reported	NA	NA

207	Jin	2017	NR	NR	Reported	NR	Reported	NR	NR
208	Li	2017	NR	Reported	Reported	NR	Reported	NA	NA
209	Maezawa	2017	NR	NR	Reported	NR	NR	NR	NR
210	Tao	2017	NR	NR	Reported	NR	Reported	NR	NR
211	Yu	2017	NR	NR	Reported	NR	Reported	NR	NR
212	Martinez Hernandez	2018	NR	NR	Reported	NR	Reported	NR	NR
213	Wang	2018	NR	NR	NR	NR	Reported	Reported	NR
214	Zhu	2018	NR	NR	Reported	NR	Reported	NR	NR
215	Gureviciene	2004	Reported	NR	Reported	NR	NR	NA	NA
216	Fitzjohn	2010	NR	Reported	NR	NR	Reported	NA	NA
217	Ricoy	2011	NR	Reported	Reported	NR	NR	NA	NA
218	Maingret	2017	NR	Reported	Reported	NR	NR	NA	NA
219	Jolas	2002	NR	Reported	NR	NR	NR	NA	NA
220	Arrieta-Cruz	2010	NR	NR	Reported	NR	NR	NA	NA
221	Arrieta-Cruz	2010	NR	NR	Reported	NR	Reported	NA	NA
222	Ye	2010	NR	NR	Reported	NR	Reported	NA	NA
223	Sclip	2011	NR	NR	Reported	NR	NR	NR	NR
224	Sclip	2014	NR	NR	NR	NR	Reported	NA	NA
225	Luccarini	2015	NR	NR	Reported	NR	Reported	NA	NA
226	Tozzi	2015	NR	NR	NR	NR	Reported	NA	NA
227	Cavanagh	2016	NR	NR	Reported	NR	NR	NR	NR
228	Hinrich	2016	NR	Reported	Reported	NR	Reported	Reported	NR
229	Kimura	2017	NR	NR	Reported	NR	Reported	NA	NA
230	Knock	2018	NR	NR	Reported	NR	NR	NA	NA
231	Oddo	2003	NR	NR	NR	NR	NR	NA	NA
232	Chakroborty	2009	NR	NR	Reported	NR	NR	NA	NA
233	Chakroborty	2012	NR	NR	Reported	NR	Reported	NA	NA
234	Chakroborty	2012	NR	NR	Reported	NR	Reported	NR	NR

235	Hsiao	2012	NR	NR	Reported	NR	NR	Reported	NR
236	Searcy	2012	Reported	NR	NR	NR	Reported	Reported	NR
237	Giannopoulos	2013	Reported	Reported	Reported	NR	Reported	Reported	NR
238	Sancheti	2013	NR	NR	Reported	NR	Reported	NR	NR
239	Giannopoulos	2014	NR	NR	Reported	NR	Reported	Reported	NR
240	Grigoryan	2014	NR	NR	Reported	NR	Reported	NR	NR
241	Shilling	2014	Reported	NR	Reported	Reported	Reported	NA	NA
242	Caccamo	2015	NR	NR	Reported	NR	Reported	NA	NA
243	Chakroborty	2015	NR	NR	NR	NR	Reported	NA	NA
244	Sykora	2015	Reported	Reported	Reported	NR	Reported	NA	NA
245	Wang	2015	NR	Reported	Reported	NR	NR	NR	Reported
246	Wang	2015	NR	NR	Reported	NR	NR	NR	NR
247	Luo	2016	NR	NR	Reported	NR	Reported	Reported	NR
248	Cisse	2017	NR	Reported	Reported	NR	Reported	NR	NR
249	Jin	2017	NR	NR	Reported	NR	NR	NR	NR
250	Sun	2017	NR	NR	Reported	NR	NR	NR	NR
251	Wang	2017	NR	NR	Reported	NR	NR	NR	NR
252	Zhao	2017	NR	Reported	Reported	NR	NR	NA	NA
253	Gelman	2018	NR	NR	NR	NR	Reported	NA	NA
254	Gong	2004	NR	Reported	Reported	NR	Reported	NR	Reported
255	Battaglia	2007	NR	NR	NR	NR	NR	NA	NA
256	Liu	2008	NR	NR	NR	NR	NR	NA	NA
257	Francis	2009	NR	Reported	Reported	NR	NR	NR	NR
258	Calella	2010	Reported	NR	NR	NR	Reported	NA	NA
259	Bachstetter	2012	NR	Reported	Reported	NR	Reported	NR	NR
260	Fiorito	2013	NR	Reported	NR	NR	NR	NR	NR
261	Rudobeck	2017	NR	Reported	Reported	NR	Reported	Reported	Reported
262	Roder	2003	NR	Reported	NR	NR	NR	NA	NA
263	Balducci	2010	Reported	NR	NR	NR	Reported	NA	NA

264	Moriguchi	2016	NR	NR	Reported	NR	Reported	NR	NR
265	Larson	1999	NR	Reported	NR	NR	NR	NA	NA
266	Hartman	2005	NR	NR	NR	NR	NR	NR	NR

NR: not reported, NA: not applicable

10. ***Data analysis code***

Available in Open Science Framework project: https://osf.io/pkwh8/

Appendix C6

1. *OFT SR protocol*

https://doi.org/10.17605/OSF.IO/E4WBF

2. *OFT regex:*

[oO]pen[-|]?[fF]ield[-|]?|OFT[|.|)|\s]|[oO]pen[-|]?[fF]eild

3. *Data extraction guide for reviewers*

https://osf.io/c5jny/

4. *Publications included in review*

Publications are listed with study ID used throughout review.

101. Halagappa, V. K., Guo, Z., Pearson, M., Matsuoka, Y., Cutler, R. G., Laferla, F. M., & Mattson, M. P. (2007). Intermittent fasting and caloric restriction ameliorate age-related behavioral deficits in the triple-transgenic mouse model of Alzheimer's disease. *Neurobiol Dis, 26*(1), 212-220. doi:10.1016/j.nbd.2006.12.019
102. Nelson, R. L., Guo, Z., Halagappa, V. M., Pearson, M., Gray, A. J., Matsuoka, Y., . . . Mattson, M. P. (2007). Prophylactic treatment with paroxetine ameliorates behavioral deficits and retards the development of amyloid and tau pathologies in 3xTgAD mice. *Exp Neurol, 205*(1), 166-176. doi:10.1016/j.expneurol.2007.01.037
103. Pietropaolo, S., Sun, Y., Li, R., Brana, C., Feldon, J., & Yee, B. K. (2008). The impact of voluntary exercise on mental health in rodents: a neuroplasticity perspective. *Behav Brain Res, 192*(1), 42-60. doi:10.1016/j.bbr.2008.03.014
104. Gulinello, M., Gertner, M., Mendoza, G., Schoenfeld, B. P., Oddo, S., LaFerla, F., . . . Faber, D. S. (2009). Validation of a 2-day water maze protocol in mice. *Behav Brain Res, 196*(2), 220-227. doi:10.1016/j.bbr.2008.09.002
105. Blanchard, J., Wanka, L., Tung, Y. C., Cardenas-Aguayo Mdel, C., LaFerla, F. M., Iqbal, K., & Grundke-Iqbal, I. (2010). Pharmacologic reversal of neurogenic and neuroplastic abnormalities and cognitive impairments without affecting Abeta and tau pathologies in 3xTg-AD mice. *Acta Neuropathol, 120*(5), 605-621. doi:10.1007/s00401-010-0734-6
106. Gimenez-Llort, L., Garcia, Y., Buccieri, K., Revilla, S., Sunol, C., Cristofol, R., & Sanfeliu, C. (2010). Gender-Specific Neuroimmunoendocrine Response to Treadmill Exercise in 3xTg-AD Mice. *Int J Alzheimers Dis, 2010*, 128354. doi:10.4061/2010/128354
107. Chadwick, W., Mitchell, N., Caroll, J., Zhou, Y., Park, S. S., Wang, L., . . . Maudsley, S. (2011). Amitriptyline-mediated cognitive enhancement in aged 3xTg Alzheimer's disease mice is associated with neurogenesis and neurotrophic activity. *PLoS One, 6*(6), e21660. doi:10.1371/journal.pone.0021660
108. Garcia-Mesa, Y., Lopez-Ramos, J. C., Gimenez-Llort, L., Revilla, S., Guerra, R., Gruart, A., . . . Sanfeliu, C. (2011). Physical exercise protects against Alzheimer's disease in 3xTg-AD mice. *J Alzheimers Dis, 24*(3), 421-454. doi:10.3233/JAD-2011-101635
109. Medeiros, R., Kitazawa, M., Caccamo, A., Baglietto-Vargas, D., Estrada-Hernandez, T., Cribbs, D. H., . . . LaFerla, F. M. (2011). Loss of muscarinic M1 receptor exacerbates Alzheimer's disease-like pathology and cognitive decline. *Am J Pathol, 179*(2), 980-991. doi:10.1016/j.ajpath.2011.04.041

110. Niikura, T., Sidahmed, E., Hirata-Fukae, C., Aisen, P. S., & Matsuoka, Y. (2011). A humanin derivative reduces amyloid beta accumulation and ameliorates memory deficit in triple transgenic mice. *PLoS One, 6*(1), e16259. doi:10.1371/journal.pone.0016259
111. Bories, C., Guitton, M. J., Julien, C., Tremblay, C., Vandal, M., Msaid, M., . . . Calon, F. (2012). Sex-dependent alterations in social behaviour and cortical synaptic activity coincide at different ages in a model of Alzheimer's disease. *PLoS One, 7*(9), e46111. doi:10.1371/journal.pone.0046111
112. Filali, M., Lalonde, R., Theriault, P., Julien, C., Calon, F., & Planel, E. (2012). Cognitive and non-cognitive behaviors in the triple transgenic mouse model of Alzheimer's disease expressing mutated APP, PS1, and Mapt (3xTg-AD). *Behav Brain Res, 234*(2), 334-342. doi:10.1016/j.bbr.2012.07.004
113. Garcia-Mesa, Y., Gimenez-Llort, L., Lopez, L. C., Venegas, C., Cristofol, R., Escames, G., . . . Sanfeliu, C. (2012). Melatonin plus physical exercise are highly neuroprotective in the 3xTg-AD mouse. *Neurobiol Aging, 33*(6), 1124 e1113-1129. doi:10.1016/j.neurobiolaging.2011.11.016
114. Rothman, S. M., Herdener, N., Camandola, S., Texel, S. J., Mughal, M. R., Cong, W. N., . . . Mattson, M. P. (2012). 3xTgAD mice exhibit altered behavior and elevated Abeta after chronic mild social stress. *Neurobiol Aging, 33*(4), 830 e831-812. doi:10.1016/j.neurobiolaging.2011.07.005
115. Carvalho, C., Machado, N., Mota, P. C., Correia, S. C., Cardoso, S., Santos, R. X., . . . Moreira, P. I. (2013). Type 2 diabetic and Alzheimer's disease mice present similar behavioral, cognitive, and vascular anomalies. *J Alzheimers Dis, 35*(3), 623-635. doi:10.3233/JAD-130005
116. Chen, Y., Liang, Z., Blanchard, J., Dai, C. L., Sun, S., Lee, M. H., . . . Gong, C. X. (2013). A non-transgenic mouse model (icv-STZ mouse) of Alzheimer's disease: similarities to and differences from the transgenic model (3xTg-AD mouse). *Mol Neurobiol, 47*(2), 711-725. doi:10.1007/s12035-012-8375-5
117. George, S., Petit, G. H., Gouras, G. K., Brundin, P., & Olsson, R. (2013). Nonsteroidal selective androgen receptor modulators and selective estrogen receptor beta agonists moderate cognitive deficits and amyloid-beta levels in a mouse model of Alzheimer's disease. *ACS Chem Neurosci, 4*(12), 1537-1548. doi:10.1021/cn400133s
118. Gimenez-Llort, L., Rivera-Hernandez, G., Marin-Argany, M., Sanchez-Quesada, J. L., & Villegas, S. (2013). Early intervention in the 3xTg-AD mice with an amyloid beta-antibody fragment ameliorates first hallmarks of Alzheimer disease. *Mabs, 5*(5), 665-677. doi:10.4161/mabs.25424
119. Hebda-Bauer, E. K., Simmons, T. A., Sugg, A., Ural, E., Stewart, J. A., Beals, J. L., . . . Akil, H. (2013). 3xTg-AD mice exhibit an activated central stress axis during early-stage pathology. *J Alzheimers Dis, 33*(2), 407-422. doi:10.3233/JAD-2012-121438
120. Liu, D., Pitta, M., Jiang, H., Lee, J. H., Zhang, G., Chen, X., . . . Mattson, M. P. (2013). Nicotinamide forestalls pathology and cognitive decline in Alzheimer mice: evidence for improved neuronal bioenergetics and autophagy procession. *Neurobiol Aging, 34*(6), 1564-1580. doi:10.1016/j.neurobiolaging.2012.11.020
121. Ratia, M., Gimenez-Llort, L., Camps, P., Munoz-Torrero, D., Perez, B., Clos, M. V., & Badia, A. (2013). Huprine X and huperzine A improve cognition and regulate some neurochemical processes related with Alzheimer's disease in triple transgenic mice (3xTg-AD). *Neurodegener Dis, 11*(3), 129-140. doi:10.1159/000336427
122. Yamamoto, S., Ooshima, Y., Nakata, M., Yano, T., Matsuoka, K., Watanabe, S., . . . Hashimoto, T. (2013). Generation of gene-targeted mice using embryonic stem cells derived from a transgenic mouse model of Alzheimer's disease. *Transgenic Res, 22*(3), 537-547. doi:10.1007/s11248-012-9651-x
123. Blázquez, G., Cañete, T., Tobeña, A., Giménez-Llort, L., & Fernández-Teruel, A. (2014). Cognitive and emotional profiles of aged Alzheimer's disease (3×TgAD) mice: Effects of environmental enrichment and sexual dimorphism. *Behavioural Brain Research, 268*, 185-201. doi:10.1016/j.bbr.2014.04.008

124. Chen, Y., Liang, Z., Tian, Z., Blanchard, J., Dai, C. L., Chalbot, S., . . . Gong, C. X. (2014). Intracerebroventricular streptozotocin exacerbates Alzheimer-like changes of 3xTg-AD mice. *Mol Neurobiol, 49*(1), 547-562. doi:10.1007/s12035-013-8539-y
125. Kazim, S. F., Blanchard, J., Dai, C. L., Tung, Y. C., LaFerla, F. M., Iqbal, I. G., & Iqbal, K. (2014). Disease modifying effect of chronic oral treatment with a neurotrophic peptidergic compound in a triple transgenic mouse model of Alzheimer's disease. *Neurobiol Dis, 71*, 110-130. doi:10.1016/j.nbd.2014.07.001
126. Pietropaolo, S., Feldon, J., & Yee, B. K. (2014). Environmental enrichment eliminates the anxiety phenotypes in a triple transgenic mouse model of Alzheimer's disease. *Cogn Affect Behav Neurosci, 14*(3), 996-1008. doi:10.3758/s13415-014-0253-3
127. St-Amour, I., Pare, I., Tremblay, C., Coulombe, K., Bazin, R., & Calon, F. (2014). IVIg protects the 3xTg-AD mouse model of Alzheimer's disease from memory deficit and Abeta pathology. *J Neuroinflammation, 11*, 54. doi:10.1186/1742-2094-11-54
128. Baeta-Corral, R., Defrin, R., Pick, C. G., & Gimenez-Llort, L. (2015). Tail-flick test response in 3xTg-AD mice at early and advanced stages of disease. *Neurosci Lett, 600*, 158-163. doi:10.1016/j.neulet.2015.06.007
129. Canete, T., Blazquez, G., Tobena, A., Gimenez-Llort, L., & Fernandez-Teruel, A. (2015). Cognitive and emotional alterations in young Alzheimer's disease (3xTgAD) mice: effects of neonatal handling stimulation and sexual dimorphism. *Behav Brain Res, 281*, 156-171. doi:10.1016/j.bbr.2014.11.004
130. Romano, A., Pace, L., Tempesta, B., Lavecchia, A. M., Macheda, T., Bedse, G., . . . Cassano, T. (2014). Depressive-like behavior is paired to monoaminergic alteration in a murine model of Alzheimer's disease. *Int J Neuropsychopharmacol, 18*(4). doi:10.1093/ijnp/pyu020
131. Torres-Lista, V., Lopez-Pousa, S., & Gimenez-Llort, L. (2015). Marble-burying is enhanced in 3xTg-AD mice, can be reversed by risperidone and it is modulable by handling. *Behav Processes, 116*, 69-74. doi:10.1016/j.beproc.2015.05.001
132. Yu, Y., Li, X., Blanchard, J., Li, Y., Iqbal, K., Liu, F., & Gong, C. X. (2015). Insulin sensitizers improve learning and attenuate tau hyperphosphorylation and neuroinflammation in 3xTg-AD mice. *J Neural Transm (Vienna), 122*(4), 593-606. doi:10.1007/s00702-014-1294-z
133. Garcia-Mesa, Y., Colie, S., Corpas, R., Cristofol, R., Comellas, F., Nebreda, A. R., . . . Sanfeliu, C. (2016). Oxidative Stress Is a Central Target for Physical Exercise Neuroprotection Against Pathological Brain Aging. *J Gerontol A Biol Sci Med Sci, 71*(1), 40-49. doi:10.1093/gerona/glv005
134. Liu, S., Tang, S. Q., Cui, H. J., Yin, S., Yin, M., Zhao, H., . . . Lu, Y. (2016). Dipotassium N-stearoyltyrosinate ameliorated pathological injuries in triple-transgenic mouse model of Alzheimer's disease. *J Pharmacol Sci, 132*(1), 92-99. doi:10.1016/j.jphs.2016.08.009
135. Magistri, M., Velmeshev, D., Makhmutova, M., Patel, P., Sartor, G. C., Volmar, C. H., . . . Faghihi, M. A. (2016). The BET-Bromodomain Inhibitor JQ1 Reduces Inflammation and Tau Phosphorylation at Ser396 in the Brain of the 3xTg Model of Alzheimer's Disease. *Curr Alzheimer Res, 13*(9), 985-995. doi:10.2174/1567205013666160427101832
136. Bonfili, L., Cecarini, V., Berardi, S., Scarpona, S., Suchodolski, J. S., Nasuti, C., . . . Eleuteri, A. M. (2017). Microbiota modulation counteracts Alzheimer's disease progression influencing neuronal proteolysis and gut hormones plasma levels. *Sci Rep, 7*(1), 2426. doi:10.1038/s41598-017-02587-2
137. Branca, C., Shaw, D. M., Belfiore, R., Gokhale, V., Shaw, A. Y., Foley, C., . . . Oddo, S. (2017). Dyrk1 inhibition improves Alzheimer's disease-like pathology. *Aging Cell, 16*(5), 1146-1154. doi:10.1111/acel.12648
138. Corpas, R., Revilla, S., Ursulet, S., Castro-Freire, M., Kaliman, P., Petegnief, V., . . . Sanfeliu, C. (2017). SIRT1 Overexpression in Mouse Hippocampus Induces Cognitive Enhancement Through Proteostatic and Neurotrophic Mechanisms. *Mol Neurobiol, 54*(7), 5604-5619. doi:10.1007/s12035-016-0087-9

139. Durairajan, S. S. K., Iyaswamy, A., Shetty, S. G., Kammella, A. K., Malampati, S., Shang, W., . . . Li, M. (2017). A modified formulation of Huanglian-Jie-Du-Tang reduces memory impairments and beta-amyloid plaques in a triple transgenic mouse model of Alzheimer's disease. *Sci Rep, 7*(1), 6238. doi:10.1038/s41598-017-06217-9
140. Esquerda-Canals, G., Marti-Clua, J., Roda, A. R., & Villegas, S. (2017). An Intracellular Amyloid-beta/AbetaPP Epitope Correlates with Neurodegeneration in those Neuronal Populations Early Involved in Alzheimer's Disease. *J Alzheimers Dis, 59*(3), 1079-1096. doi:10.3233/JAD-170218
141. Hussain, R. Z., Miller-Little, W. A., Lambracht-Washington, D., Jaramillo, T. C., Takahashi, M., Zhang, S., . . . Stuve, O. (2017). Laquinimod has no effects on brain volume or cellular CNS composition in the F1 3xTg-AD/C3H mouse model of Alzheimer's disease. *J Neuroimmunol, 309*, 100-110. doi:10.1016/j.jneuroim.2017.05.017
142. Nie, L., Wei, G., Peng, S., Qu, Z., Yang, Y., Yang, Q., . . . Yang, X. (2017). Melatonin ameliorates anxiety and depression-like behaviors and modulates proteomic changes in triple transgenic mice of Alzheimer's disease. *Biofactors, 43*(4), 593-611. doi:10.1002/biof.1369
143. Nie, L., Xia, J., Li, H., Zhang, Z., Yang, Y., Huang, X., . . . Yang, X. (2017). Ginsenoside Rg1 Ameliorates Behavioral Abnormalities and Modulates the Hippocampal Proteomic Change in Triple Transgenic Mice of Alzheimer's Disease. *Oxid Med Cell Longev, 2017*, 6473506. doi:10.1155/2017/6473506
144. Volmar, C. H., Salah-Uddin, H., Janczura, K. J., Halley, P., Lambert, G., Wodrich, A., . . . Wahlestedt, C. (2017). M344 promotes nonamyloidogenic amyloid precursor protein processing while normalizing Alzheimer's disease genes and improving memory. *Proc Natl Acad Sci U S A, 114*(43), E9135-E9144. doi:10.1073/pnas.1707544114
145. Joyashiki, E., Matsuya, Y., & Tohda, C. (2011). Sominone improves memory impairments and increases axonal density in Alzheimer's disease model mice, 5XFAD. *Int J Neurosci, 121*(4), 181-190. doi:10.3109/00207454.2010.541571
146. Hillmann, A., Hahn, S., Schilling, S., Hoffmann, T., Demuth, H. U., Bulic, B., . . . Wirths, O. (2012). No improvement after chronic ibuprofen treatment in the 5XFAD mouse model of Alzheimer's disease. *Neurobiol Aging, 33*(4), 833 e839-850. doi:10.1016/j.neurobiolaging.2011.08.006
147. Corbett, G. T., Roy, A., & Pahan, K. (2013). Sodium phenylbutyrate enhances astrocytic neurotrophin synthesis via protein kinase C (PKC)-mediated activation of cAMP-response element-binding protein (CREB): implications for Alzheimer disease therapy. *J Biol Chem, 288*(12), 8299-8312. doi:10.1074/jbc.M112.426536
148. Bhattacharya, S., Haertel, C., Maelicke, A., & Montag, D. (2014). Galantamine slows down plaque formation and behavioral decline in the 5XFAD mouse model of Alzheimer's disease. *PLoS One, 9*(2), e89454. doi:10.1371/journal.pone.0089454
149. Flanigan, T. J., Xue, Y., Kishan Rao, S., Dhanushkodi, A., & McDonald, M. P. (2014). Abnormal vibrissa-related behavior and loss of barrel field inhibitory neurons in 5xFAD transgenics. *Genes Brain Behav, 13*(5), 488-500. doi:10.1111/gbb.12133
150. Schneider, F., Baldauf, K., Wetzel, W., & Reymann, K. G. (2014). Behavioral and EEG changes in male 5xFAD mice. *Physiol Behav, 135*, 25-33. doi:10.1016/j.physbeh.2014.05.041
151. Zhang, J., Hu, M., Teng, Z., Tang, Y. P., & Chen, C. (2014). Synaptic and cognitive improvements by inhibition of 2-AG metabolism are through upregulation of microRNA-188-3p in a mouse model of Alzheimer's disease. *J Neurosci, 34*(45), 14919-14933. doi:10.1523/JNEUROSCI.1165-14.2014
152. Bhattacharya, S., Maelicke, A., & Montag, D. (2015). Nasal Application of the Galantamine Pro-drug Memogain Slows Down Plaque Deposition and Ameliorates Behavior in 5X Familial Alzheimer's Disease Mice. *J Alzheimers Dis, 46*(1), 123-136. doi:10.3233/JAD-142421
153. J., J. Y. J. K. G. Y. K. J. H. C. H. D. P. J. K. K. N. L. Y. S. a. L. H. (2015). 1950 MHz Electromagnetic Fields Ameliorate A beta Pathology in Alzheimer's Disease Mice. *Current Alzheimer Research, 12*(5), 481-492. doi:NA

154. Paesler, K., Xie, K., Hettich, M. M., Siwek, M. E., Ryan, D. P., Schroder, S., . . . Ehninger, D. (2015). Limited effects of an eIF2alphaS51A allele on neurological impairments in the 5xFAD mouse model of Alzheimer's disease. *Neural Plast, 2015*, 825157. doi:10.1155/2015/825157
155. Schneider, F., Baldauf, K., Wetzel, W., & Reymann, K. G. (2015). Effects of methylphenidate on the behavior of male 5xFAD mice. *Pharmacol Biochem Behav, 128*, 68-77. doi:10.1016/j.pbb.2014.11.006
156. Woo, J. A., Boggess, T., Uhlar, C., Wang, X., Khan, H., Cappos, G., . . . Kang, D. E. (2015). RanBP9 at the intersection between cofilin and Abeta pathologies: rescue of neurodegenerative changes by RanBP9 reduction. *Cell Death Dis, 6*, 1676. doi:10.1038/cddis.2015.37
157. Grinan-Ferre, C., Sarroca, S., Ivanova, A., Puigoriol-Illamola, D., Aguado, F., Camins, A., . . . Pallas, M. (2016). Epigenetic mechanisms underlying cognitive impairment and Alzheimer disease hallmarks in 5XFAD mice. *Aging (Albany NY), 8*(4), 664-684. doi:10.18632/aging.100906
158. Nikolaeva, N. S., Sokolov, V. B., Aksinenko, A. Y., Ovchinnikov, R. K., Bachurin, S. O., & Kinzirskii, A. S. (2016). Psychotropic Activity of New Fluorinated Derivatives of Tetrahydrocarbasoles. *Bull Exp Biol Med, 160*(4), 455-458. doi:10.1007/s10517-016-3195-7
159. Son, Y., Jeong, Y. J., Kwon, J. H., Choi, H. D., Pack, J. K., Kim, N., . . . Lee, H. J. (2016). 1950 MHz radiofrequency electromagnetic fields do not aggravate memory deficits in 5xFAD mice. *Bioelectromagnetics, 37*(6), 391-399. doi:10.1002/bem.21992
160. Tang, X., Wu, D., Gu, L. H., Nie, B. B., Qi, X. Y., Wang, Y. J., . . . Zhang, Z. J. (2016). Spatial learning and memory impairments are associated with increased neuronal activity in 5XFAD mouse as measured by manganese-enhanced magnetic resonance imaging. *Oncotarget, 7*(36), 57556-57570. doi:10.18632/oncotarget.11353
161. Ardestani, P. M., Evans, A. K., Yi, B., Nguyen, T., Coutellier, L., & Shamloo, M. (2017). Modulation of neuroinflammation and pathology in the 5XFAD mouse model of Alzheimer's disease using a biased and selective beta-1 adrenergic receptor partial agonist. *Neuropharmacology, 116*, 371-386. doi:10.1016/j.neuropharm.2017.01.010
162. Baranger, K., Bonnet, A. E., Girard, S. D., Paumier, J. M., Garcia-Gonzalez, L., Elmanaa, W., . . . Rivera, S. (2016). MT5-MMP Promotes Alzheimer's Pathogenesis in the Frontal Cortex of 5xFAD Mice and APP Trafficking in vitro. *Front Mol Neurosci, 9*, 163. doi:10.3389/fnmol.2016.00163
163. Brandscheid, C., Schuck, F., Reinhardt, S., Schafer, K. H., Pietrzik, C. U., Grimm, M., . . . Endres, K. (2017). Altered Gut Microbiome Composition and Tryptic Activity of the 5xFAD Alzheimer's Mouse Model. *J Alzheimers Dis, 56*(2), 775-788. doi:10.3233/JAD-160926
164. Braun, D., & Feinstein, D. L. (2019). The locus coeruleus neuroprotective drug vindeburnol normalizes behavior in the 5xFAD transgenic mouse model of Alzheimer's disease. *Brain Res, 1702*, 29-37. doi:10.1016/j.brainres.2017.12.028
165. Nakagawa, T., Hasegawa, Y., Uekawa, K., Senju, S., Nakagata, N., Matsui, K., & Kim-Mitsuyama, S. (2017). Transient Mild Cerebral Ischemia Significantly Deteriorated Cognitive Impairment in a Mouse Model of Alzheimer's Disease via Angiotensin AT1 Receptor. *Am J Hypertens, 30*(2), 141-150. doi:10.1093/ajh/hpw099
166. Sawmiller, D., Li, S., Mori, T., Habib, A., Rongo, D., Delic, V., . . . Tan, J. (2017). Beneficial effects of a pyrroloquinolinequinone-containing dietary formulation on motor deficiency, cognitive decline and mitochondrial dysfunction in a mouse model of Alzheimer's disease. *Heliyon, 3*(4), e00279. doi:10.1016/j.heliyon.2017.e00279
167. Wu, P. J., Hung, Y. F., Liu, H. Y., & Hsueh, Y. P. (2017). Deletion of the Inflammasome Sensor Aim2 Mitigates Abeta Deposition and Microglial Activation but Increases Inflammatory Cytokine Expression in an Alzheimer Disease Mouse Model. *Neuroimmunomodulation, 24*(1), 29-39. doi:10.1159/000477092
168. Yang, E. J., Mahmood, U., Kim, H., Choi, M., Choi, Y., Lee, J. P., . . . Kim, H. S. (2017). Alterations in protein phosphorylation in the amygdala of the 5XFamilial Alzheimer's disease animal model. *J Pharmacol Sci, 133*(4), 261-267. doi:10.1016/j.jphs.2017.03.005

169. Zhen, J., Qian, Y., Weng, X., Su, W., Zhang, J., Cai, L., . . . Wang, X. (2017). Gamma rhythm low field magnetic stimulation alleviates neuropathologic changes and rescues memory and cognitive impairments in a mouse model of Alzheimer's disease. *Alzheimers Dement (N Y), 3*(4), 487-497. doi:10.1016/j.trci.2017.07.002
170. He, Z., Guo, J. L., McBride, J. D., Narasimhan, S., Kim, H., Changolkar, L., . . . Lee, V. M. (2018). Amyloid-beta plaques enhance Alzheimer's brain tau-seeded pathologies by facilitating neuritic plaque tau aggregation. *Nat Med, 24*(1), 29-38. doi:10.1038/nm.4443
171. O'Leary, T. P., Robertson, A., Chipman, P. H., Rafuse, V. F., & Brown, R. E. (2018). Motor function deficits in the 12 month-old female 5xFAD mouse model of Alzheimer's disease. *Behav Brain Res, 337*, 256-263. doi:10.1016/j.bbr.2017.09.009
172. Son, Y., Kim, J. S., Jeong, Y. J., Jeong, Y. K., Kwon, J. H., Choi, H. D., . . . Lee, H. J. (2018). Long-term RF exposure on behavior and cerebral glucose metabolism in 5xFAD mice. *Neurosci Lett, 666*, 64-69. doi:10.1016/j.neulet.2017.12.042
173. Van Dam, D., D'Hooge, R., Staufenbiel, M., Van Ginneken, C., Van Meir, F., & De Deyn, P. P. (2003). Age-dependent cognitive decline in the APP23 model precedes amyloid deposition. *Eur J Neurosci, 17*(2), 388-396. doi:10.1046/j.1460-9568.2003.02444.x
174. Dumont, M., Strazielle, C., Staufenbiel, M., & Lalonde, R. (2004). Spatial learning and exploration of environmental stimuli in 24-month-old female APP23 transgenic mice with the Swedish mutation. *Brain Res, 1024*(1-2), 113-121. doi:10.1016/j.brainres.2004.07.052
175. Lalonde, R., Dumont, M., Staufenbiel, M., & Strazielle, C. (2005). Neurobehavioral characterization of APP23 transgenic mice with the SHIRPA primary screen. *Behav Brain Res, 157*(1), 91-98. doi:10.1016/j.bbr.2004.06.020
176. Heneka, M. T., Ramanathan, M., Jacobs, A. H., Dumitrescu-Ozimek, L., Bilkei-Gorzo, A., Debeir, T., . . . Staufenbiel, M. (2006). Locus ceruleus degeneration promotes Alzheimer pathogenesis in amyloid precursor protein 23 transgenic mice. *J Neurosci, 26*(5), 1343-1354. doi:10.1523/JNEUROSCI.4236-05.2006
177. Arendash, G. W., Jensen, M. T., Salem, N., Jr., Hussein, N., Cracchiolo, J., Dickson, A., . . . Potter, H. (2007). A diet high in omega-3 fatty acids does not improve or protect cognitive performance in Alzheimer's transgenic mice. *Neuroscience, 149*(2), 286-302. doi:10.1016/j.neuroscience.2007.08.018
178. Hernandez, C. M., Kayed, R., Zheng, H., Sweatt, J. D., & Dineley, K. T. (2010). Loss of alpha7 nicotinic receptors enhances beta-amyloid oligomer accumulation, exacerbating early-stage cognitive decline and septohippocampal pathology in a mouse model of Alzheimer's disease. *J Neurosci, 30*(7), 2442-2453. doi:10.1523/JNEUROSCI.5038-09.2010
179. Terwel, D., Steffensen, K. R., Verghese, P. B., Kummer, M. P., Gustafsson, J. A., Holtzman, D. M., & Heneka, M. T. (2011). Critical role of astroglial apolipoprotein E and liver X receptor-alpha expression for microglial Abeta phagocytosis. *J Neurosci, 31*(19), 7049-7059. doi:10.1523/JNEUROSCI.6546-10.2011
180. Katsouri, L., Lim, Y. M., Blondrath, K., Eleftheriadou, I., Lombardero, L., Birch, A. M., . . . Sastre, M. (2016). PPARgamma-coactivator-1alpha gene transfer reduces neuronal loss and amyloid-beta generation by reducing beta-secretase in an Alzheimer's disease model. *Proc Natl Acad Sci U S A, 113*(43), 12292-12297. doi:10.1073/pnas.1606171113
181. Dineley, K. T., Xia, X., Bui, D., Sweatt, J. D., & Zheng, H. (2002). Accelerated plaque accumulation, associative learning deficits, and up-regulation of alpha 7 nicotinic receptor protein in transgenic mice co-expressing mutant human presenilin 1 and amyloid precursor proteins. *J Biol Chem, 277*(25), 22768-22780. doi:10.1074/jbc.M200164200
182. Liu, L., Orozco, I. J., Planel, E., Wen, Y., Bretteville, A., Krishnamurthy, P., . . . Duff, K. (2008). A transgenic rat that develops Alzheimer's disease-like amyloid pathology, deficits in synaptic plasticity and cognitive impairment. *Neurobiol Dis, 31*(1), 46-57. doi:10.1016/j.nbd.2008.03.005
183. Jardanhazi-Kurutz, D., Kummer, M. P., Terwel, D., Vogel, K., Dyrks, T., Thiele, A., & Heneka, M. T. (2010). Induced LC degeneration in APP/PS1 transgenic mice accelerates early cerebral

amyloidosis and cognitive deficits. *Neurochem Int, 57*(4), 375-382. doi:10.1016/j.neuint.2010.02.001
184. Govindarajan, N., Agis-Balboa, R. C., Walter, J., Sananbenesi, F., & Fischer, A. (2011). Sodium butyrate improves memory function in an Alzheimer's disease mouse model when administered at an advanced stage of disease progression. *J Alzheimers Dis, 26*(1), 187-197. doi:10.3233/JAD-2011-110080
185. Vom Berg, J., Prokop, S., Miller, K. R., Obst, J., Kalin, R. E., Lopategui-Cabezas, I., . . . Heppner, F. L. (2012). Inhibition of IL-12/IL-23 signaling reduces Alzheimer's disease-like pathology and cognitive decline. *Nat Med, 18*(12), 1812-1819. doi:10.1038/nm.2965
186. Agis-Balboa, R. C., Pavelka, Z., Kerimoglu, C., & Fischer, A. (2013). Loss of HDAC5 impairs memory function: implications for Alzheimer's disease. *J Alzheimers Dis, 33*(1), 35-44. doi:10.3233/JAD-2012-121009
187. Ferguson, S. A., Sarkar, S., & Schmued, L. C. (2013). Longitudinal behavioral changes in the APP/PS1 transgenic Alzheimer's disease model. *Behav Brain Res, 242*, 125-134. doi:10.1016/j.bbr.2012.12.055
188. Gao, L., Tian, S., Gao, H., & Xu, Y. (2013). Hypoxia increases Abeta-induced tau phosphorylation by calpain and promotes behavioral consequences in AD transgenic mice. *J Mol Neurosci, 51*(1), 138-147. doi:10.1007/s12031-013-9966-y
189. Govindarajan, N., Rao, P., Burkhardt, S., Sananbenesi, F., Schluter, O. M., Bradke, F., . . . Fischer, A. (2013). Reducing HDAC6 ameliorates cognitive deficits in a mouse model for Alzheimer's disease. *EMBO Mol Med, 5*(1), 52-63. doi:10.1002/emmm.201201923
190. Lo, A. C., Iscru, E., Blum, D., Tesseur, I., Callaerts-Vegh, Z., Buee, L., . . . D'Hooge, R. (2013). Amyloid and tau neuropathology differentially affect prefrontal synaptic plasticity and cognitive performance in mouse models of Alzheimer's disease. *J Alzheimers Dis, 37*(1), 109-125. doi:10.3233/JAD-122296
191. Kummer, M. P., Schwarzenberger, R., Sayah-Jeanne, S., Dubernet, M., Walczak, R., Hum, D. W., . . . Heneka, M. T. (2015). Pan-PPAR modulation effectively protects APP/PS1 mice from amyloid deposition and cognitive deficits. *Mol Neurobiol, 51*(2), 661-671. doi:10.1007/s12035-014-8743-4
192. Psotta, L., Rockahr, C., Gruss, M., Kirches, E., Braun, K., Lessmann, V., . . . Endres, T. (2015). Impact of an additional chronic BDNF reduction on learning performance in an Alzheimer mouse model. *Front Behav Neurosci, 9*, 58. doi:10.3389/fnbeh.2015.00058
193. Woo, J. A., Zhao, X., Khan, H., Penn, C., Wang, X., Joly-Amado, A., . . . Kang, D. E. (2015). Slingshot-Cofilin activation mediates mitochondrial and synaptic dysfunction via Abeta ligation to beta1-integrin conformers. *Cell Death Differ, 22*(6), 921-934. doi:10.1038/cdd.2015.5
194. Zhu, S., Shi, R., Li, V., Wang, J., Zhang, R., Tempier, A., . . . Li, X. M. (2014). Quetiapine attenuates glial activation and proinflammatory cytokines in APP/PS1 transgenic mice via inhibition of nuclear factor-kappaB pathway. *Int J Neuropsychopharmacol, 18*(3). doi:10.1093/ijnp/pyu022
195. Zhou, D., Liu, H., Li, C., Wang, F., Shi, Y., Liu, L., . . . Chen, Z. (2016). Atorvastatin ameliorates cognitive impairment, Abeta1-42 production and Tau hyperphosphorylation in APP/PS1 transgenic mice. *Metab Brain Dis, 31*(3), 693-703. doi:10.1007/s11011-016-9803-4
196. Cifuentes, D., Poittevin, M., Bonnin, P., Ngkelo, A., Kubis, N., Merkulova-Rainon, T., & Levy, B. I. (2017). Inactivation of Nitric Oxide Synthesis Exacerbates the Development of Alzheimer Disease Pathology in APPPS1 Mice (Amyloid Precursor Protein/Presenilin-1). *Hypertension, 70*(3), 613-623. doi:10.1161/HYPERTENSIONAHA.117.09742
197. Du, X., Huo, X., Yang, Y., Hu, Z., Botchway, B. O. A., Jiang, Y., & Fang, M. (2017). miR-124 downregulates BACE 1 and alters autophagy in APP/PS1 transgenic mice. *Toxicol Lett, 280*, 195-205. doi:10.1016/j.toxlet.2017.08.082
198. Geng, J., Liu, W., Xiong, Y., Ding, H., Jiang, C., Yang, X., . . . Gao, J. (2018). Andrographolide sulfonate improves Alzheimer-associated phenotypes and mitochondrial dysfunction in

APP/PS1 transgenic mice. *Biomed Pharmacother, 97*, 1032-1039. doi:10.1016/j.biopha.2017.11.039
199. Kelly, P., Denver, P., Satchell, S. C., Ackermann, M., Konerding, M. A., & Mitchell, C. A. (2017). Microvascular ultrastructural changes precede cognitive impairment in the murine APPswe/PS1dE9 model of Alzheimer's disease. *Angiogenesis, 20*(4), 567-580. doi:10.1007/s10456-017-9568-3
200. Liu, Y., Xu, Y., Zhang, L., Huang, L., Yu, P., Zhu, H., . . . Qin, C. (2017). Down-Regulated Drebrin Aggravates Cognitive Impairments in a Mouse Model of Alzheimer's Disease. *Int J Mol Sci, 18*(4). doi:10.3390/ijms18040800
201. Wu, M., Shi, H., He, Y., Yuan, L., Qu, X., Zhang, J., . . . Qi, J. (2017). Colivelin Ameliorates Impairments in Cognitive Behaviors and Synaptic Plasticity in APP/PS1 Transgenic Mice. *J Alzheimers Dis, 59*(3), 1067-1078. doi:10.3233/JAD-170307
202. Zhang, J., Zhang, R., Zhan, Z., Li, X., Zhou, F., Xing, A., . . . An, L. (2017). Beneficial Effects of Sulforaphane Treatment in Alzheimer's Disease May Be Mediated through Reduced HDAC1/3 and Increased P75NTR Expression. *Front Aging Neurosci, 9*, 121. doi:10.3389/fnagi.2017.00121
203. Wang, C. M., Devries, S., Camboni, M., Glass, M., & Martin, P. T. (2010). Immunization with the SDPM1 peptide lowers amyloid plaque burden and improves cognitive function in the APPswePSEN1(A246E) transgenic mouse model of Alzheimer's disease. *Neurobiol Dis, 39*(3), 409-422. doi:10.1016/j.nbd.2010.05.013
204. Ke, H. C., Huang, H. J., Liang, K. C., & Hsieh-Li, H. M. (2011). Selective improvement of cognitive function in adult and aged APP/PS1 transgenic mice by continuous non-shock treadmill exercise. *Brain Res, 1403*, 1-11. doi:10.1016/j.brainres.2011.05.056
205. Filali, M., & Lalonde, R. (2013). The effects of subchronic d-serine on left-right discrimination learning, social interaction, and exploratory activity in APPswe/PS1 mice. *Eur J Pharmacol, 701*(1-3), 152-158. doi:10.1016/j.ejphar.2012.12.018
206. Zhang, J., Guo, J., Zhao, X., Chen, Z., Wang, G., Liu, A., . . . Wang, C. (2013). Phosphodiesterase-5 inhibitor sildenafil prevents neuroinflammation, lowers beta-amyloid levels and improves cognitive performance in APP/PS1 transgenic mice. *Behav Brain Res, 250*, 230-237. doi:10.1016/j.bbr.2013.05.017
207. Filali, M., & Lalonde, R. (2015). Motor activity in young APPswe + PS1/A246E bigenic mice as a predicting variable for memory decline. *J Neurosci Res, 93*(6), 948-953. doi:10.1002/jnr.23552
208. Huang, H., Nie, S., Cao, M., Marshall, C., Gao, J., Xiao, N., . . . Xiao, M. (2016). Characterization of AD-like phenotype in aged APPSwe/PS1dE9 mice. *Age (Dordr), 38*(4), 303-322. doi:10.1007/s11357-016-9929-7
209. Wei, T., Yi, M., Gu, W., Hou, L., Lu, Q., Yu, Z., & Chen, H. (2016). The Potassium Channel KCa3.1 Represents a Valid Pharmacological Target for Astrogliosis-Induced Neuronal Impairment in a Mouse Model of Alzheimer's Disease. *Front Pharmacol, 7*, 528. doi:10.3389/fphar.2016.00528
210. Lalonde, R., Kim, H. D., Maxwell, J. A., & Fukuchi, K. (2005). Exploratory activity and spatial learning in 12-month-old APP(695)SWE/co+PS1/DeltaE9 mice with amyloid plaques. *Neurosci Lett, 390*(2), 87-92. doi:10.1016/j.neulet.2005.08.028
211. Frye, C. A., & Walf, A. A. (2008). Effects of progesterone administration and APPswe+PSEN1Deltae9 mutation for cognitive performance of mid-aged mice. *Neurobiol Learn Mem, 89*(1), 17-26. doi:10.1016/j.nlm.2007.09.008
212. Liu, Y., Yoo, M. J., Savonenko, A., Stirling, W., Price, D. L., Borchelt, D. R., . . . Lee, M. K. (2008). Amyloid pathology is associated with progressive monoaminergic neurodegeneration in a transgenic mouse model of Alzheimer's disease. *J Neurosci, 28*(51), 13805-13814. doi:10.1523/JNEUROSCI.4218-08.2008
213. Frye, C. A., & Walf, A. A. (2009). Progesterone reduces depression-like behavior in a murine model of Alzheimer's Disease. *Age (Dordr), 31*(2), 143-153. doi:10.1007/s11357-009-9091-6
214. Hartmann, J., Kiewert, C., & Klein, J. (2010). Neurotransmitters and energy metabolites in amyloid-bearing APP(SWE)xPSEN1dE9 Mouse Brain. *J Pharmacol Exp Ther, 332*(2), 364-370. doi:10.1124/jpet.109.161091

215. Park, S.-W., Ko, H.-G., Lee, N., Lee, H.-R., Rim, Y.-S., Kim, H., . . . Kaang, B.-K. (2010). Aged wild-type littermates and APPswe+PS1/ΔE9 mice present similar deficits in associative learning and spatial memory independent of amyloid load. *Genes & Genomics, 32*(1), 63-70. doi:10.1007/s13258-010-0833-8
216. Bonardi, C., de Pulford, F., Jennings, D., & Pardon, M. C. (2011). A detailed analysis of the early context extinction deficits seen in APPswe/PS1dE9 female mice and their relevance to preclinical Alzheimer's disease. *Behav Brain Res, 222*(1), 89-97. doi:10.1016/j.bbr.2011.03.041
217. Butler, D., Hwang, J., Estick, C., Nishiyama, A., Kumar, S. S., Baveghems, C., . . . Bahr, B. A. (2011). Protective effects of positive lysosomal modulation in Alzheimer's disease transgenic mouse models. *PLoS One, 6*(6), e20501. doi:10.1371/journal.pone.0020501
218. Tamayev, R., Matsuda, S., Giliberto, L., Arancio, O., & D'Adamio, L. (2011). APP heterozygosity averts memory deficit in knockin mice expressing the Danish dementia BRI2 mutant. *Embo j, 30*(12), 2501-2509. doi:10.1038/emboj.2011.161
219. Lim, J. E., Song, M., Jin, J., Kou, J., Pattanayak, A., Lalonde, R., & Fukuchi, K. (2012). The effects of MyD88 deficiency on exploratory activity, anxiety, motor coordination, and spatial learning in C57BL/6 and APPswe/PS1dE9 mice. *Behav Brain Res, 227*(1), 36-42. doi:10.1016/j.bbr.2011.10.027
220. Wang, D., Gao, K., Li, X., Shen, X., Zhang, X., Ma, C., . . . Zhang, L. (2012). Long-term naringin consumption reverses a glucose uptake defect and improves cognitive deficits in a mouse model of Alzheimer's disease. *Pharmacol Biochem Behav, 102*(1), 13-20. doi:10.1016/j.pbb.2012.03.013
221. Hammerschmidt, T., Kummer, M. P., Terwel, D., Martinez, A., Gorji, A., Pape, H. C., . . . Heneka, M. T. (2013). Selective loss of noradrenaline exacerbates early cognitive dysfunction and synaptic deficits in APP/PS1 mice. *Biol Psychiatry, 73*(5), 454-463. doi:10.1016/j.biopsych.2012.06.013
222. Heneka, M. T., Kummer, M. P., Stutz, A., Delekate, A., Schwartz, S., Vieira-Saecker, A., . . . Golenbock, D. T. (2013). NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature, 493*(7434), 674-678. doi:10.1038/nature11729
223. Jansen, D., Zerbi, V., Arnoldussen, I. A., Wiesmann, M., Rijpma, A., Fang, X. T., . . . Kiliaan, A. J. (2013). Effects of specific multi-nutrient enriched diets on cerebral metabolism, cognition and neuropathology in AbetaPPswe-PS1dE9 mice. *PLoS One, 8*(9), e75393. doi:10.1371/journal.pone.0075393
224. Jansen, D., Zerbi, V., Janssen, C. I., Dederen, P. J., Mutsaers, M. P., Hafkemeijer, A., . . . Kiliaan, A. J. (2013). A longitudinal study of cognition, proton MR spectroscopy and synaptic and neuronal pathology in aging wild-type and AbetaPPswe-PS1dE9 mice. *PLoS One, 8*(5), e63643. doi:10.1371/journal.pone.0063643
225. Lok, K., Zhao, H., Shen, H., Wang, Z., Gao, X., Zhao, W., & Yin, M. (2013). Characterization of the APP/PS1 mouse model of Alzheimer's disease in senescence accelerated background. *Neurosci Lett, 557 Pt B*, 84-89. doi:10.1016/j.neulet.2013.10.051
226. Lok, K., Zhao, H., Zhang, C., He, N., Shen, H., Wang, Z., . . . Yin, M. (2013). Effects of accelerated senescence on learning and memory, locomotion and anxiety-like behavior in APP/PS1 mouse model of Alzheimer's disease. *J Neurol Sci, 335*(1-2), 145-154. doi:10.1016/j.jns.2013.09.018
227. Ramos-Rodriguez, J. J., Molina-Gil, S., Rey-Brea, R., Berrocoso, E., & Garcia-Alloza, M. (2013). Specific serotonergic denervation affects tau pathology and cognition without altering senile plaques deposition in APP/PS1 mice. *PLoS One, 8*(11), e79947. doi:10.1371/journal.pone.0079947
228. Zhang, L., Ma, Q., Yang, W., Qi, X., Yao, Z., Liu, Y., . . . Qin, C. (2013). Recombinant DNA vaccine against neurite outgrowth inhibitors attenuates behavioral deficits and decreases Abeta in an Alzheimer's disease mouse model. *Neuropharmacology, 70*, 200-210. doi:10.1016/j.neuropharm.2012.10.023
229. Bernstein, K. E., Koronyo, Y., Salumbides, B. C., Sheyn, J., Pelissier, L., Lopes, D. H., . . . Koronyo-Hamaoui, M. (2014). Angiotensin-converting enzyme overexpression in myelomonocytes

prevents Alzheimer's-like cognitive decline. *J Clin Invest, 124*(3), 1000-1012. doi:10.1172/JCI66541
230. Cheng, S., LeBlanc, K. J., & Li, L. (2014). Triptolide preserves cognitive function and reduces neuropathology in a mouse model of Alzheimer's disease. *PLoS One, 9*(9), e108845. doi:10.1371/journal.pone.0108845
231. Hamilton, A., Esseltine, J. L., DeVries, R. A., Cregan, S. P., & Ferguson, S. S. (2014). Metabotropic glutamate receptor 5 knockout reduces cognitive impairment and pathogenesis in a mouse model of Alzheimer's disease. *Mol Brain, 7*, 40. doi:10.1186/1756-6606-7-40
232. Hong, X., Liu, J., Zhu, G., Zhuang, Y., Suo, H., Wang, P., . . . Huang, F. (2014). Parkin overexpression ameliorates hippocampal long-term potentiation and beta-amyloid load in an Alzheimer's disease mouse model. *Hum Mol Genet, 23*(4), 1056-1072. doi:10.1093/hmg/ddt501
233. Hsiao, Y. H., Hung, H. C., Chen, S. H., & Gean, P. W. (2014). Social interaction rescues memory deficit in an animal model of Alzheimer's disease by increasing BDNF-dependent hippocampal neurogenesis. *J Neurosci, 34*(49), 16207-16219. doi:10.1523/JNEUROSCI.0747-14.2014
234. Maroof, N., Ravipati, S., Pardon, M. C., Barrett, D. A., & Kendall, D. A. (2014). Reductions in endocannabinoid levels and enhanced coupling of cannabinoid receptors in the striatum are accompanied by cognitive impairments in the AbetaPPswe/PS1DeltaE9 mouse model of Alzheimer's disease. *J Alzheimers Dis, 42*(1), 227-245. doi:10.3233/JAD-131961
235. Wang, D., Liu, L., Zhu, X., Wu, W., & Wang, Y. (2014). Hesperidin alleviates cognitive impairment, mitochondrial dysfunction and oxidative stress in a mouse model of Alzheimer's disease. *Cell Mol Neurobiol, 34*(8), 1209-1221. doi:10.1007/s10571-014-0098-x
236. Zhang, L., Liu, C., Wu, J., Tao, J. J., Sui, X. L., Yao, Z. G., . . . Qin, C. (2014). Tubastatin A/ACY-1215 improves cognition in Alzheimer's disease transgenic mice. *J Alzheimers Dis, 41*(4), 1193-1205. doi:10.3233/JAD-140066
237. Akhter, H., Ballinger, C., Liu, N., van Groen, T., Postlethwait, E. M., & Liu, R. M. (2015). Cyclic Ozone Exposure Induces Gender-Dependent Neuropathology and Memory Decline in an Animal Model of Alzheimer's Disease. *Toxicol Sci, 147*(1), 222-234. doi:10.1093/toxsci/kfv124
238. Frost, J. L., Liu, B., Rahfeld, J. U., Kleinschmidt, M., O'Nuallain, B., Le, K. X., . . . Lemere, C. A. (2015). An anti-pyroglutamate-3 Abeta vaccine reduces plaques and improves cognition in APPswe/PS1DeltaE9 mice. *Neurobiol Aging, 36*(12), 3187-3199. doi:10.1016/j.neurobiolaging.2015.08.021
239. Huang, H., Wang, L., Cao, M., Marshall, C., Gao, J., Xiao, N., . . . Xiao, M. (2015). Isolation Housing Exacerbates Alzheimer's Disease-Like Pathophysiology in Aged APP/PS1 Mice. *Int J Neuropsychopharmacol, 18*(7), pyu116. doi:10.1093/ijnp/pyu116
240. O'Neal-Moffitt, G., Delic, V., Bradshaw, P. C., & Olcese, J. (2015). Prophylactic melatonin significantly reduces Alzheimer's neuropathology and associated cognitive deficits independent of antioxidant pathways in AbetaPP(swe)/PS1 mice. *Mol Neurodegener, 10*(25), 27. doi:10.1186/s13024-015-0027-6
241. C., W. F. S. X. L. S. C. L. W. Y. Q. J. Z. G. P. Y. F. X. L. R. a. L. (2015). Splenocytes derived from young WT mice prevent AD progression in APPswe/PSENIdE9 transgenic mice. *Oncotarget, 6*(25), 20851-20862. doi:NA
242. Li, X., Zhu, H., Sun, X., Zuo, F., Lei, J., Wang, Z., . . . Wang, R. (2016). Human Neural Stem Cell Transplantation Rescues Cognitive Defects in APP/PS1 Model of Alzheimer's Disease by Enhancing Neuronal Connectivity and Metabolic Activity. *Front Aging Neurosci, 8*, 282. doi:10.3389/fnagi.2016.00282
243. Manocha, G. D., Floden, A. M., Rausch, K., Kulas, J. A., McGregor, B. A., Rojanathammanee, L., . . . Combs, C. K. (2016). APP Regulates Microglial Phenotype in a Mouse Model of Alzheimer's Disease. *J Neurosci, 36*(32), 8471-8486. doi:10.1523/JNEUROSCI.4654-15.2016
244. Mao, Y. F., Guo, Z., Zheng, T., Jiang, Y., Yan, Y., Yin, X., . . . Zhang, B. (2016). Intranasal insulin alleviates cognitive deficits and amyloid pathology in young adult APPswe/PS1dE9 mice. *Aging Cell, 15*(5), 893-902. doi:10.1111/acel.12498

245. Mazzitelli, S., Filipello, F., Rasile, M., Lauranzano, E., Starvaggi-Cucuzza, C., Tamborini, M., . . . Matteoli, M. (2016). Amyloid-beta 1-24 C-terminal truncated fragment promotes amyloid-beta 1-42 aggregate formation in the healthy brain. *Acta Neuropathol Commun, 4*(1), 110. doi:10.1186/s40478-016-0381-9
246. Olesen, L. O., Bouzinova, E. V., Severino, M., Sivasaravanaparan, M., Hasselstrom, J. B., Finsen, B., & Wiborg, O. (2016). Behavioural Phenotyping of APPswe/PS1deltaE9 Mice: Age-Rrelated Changes and Effect of Long-Term Paroxetine Treatment. *PLoS One, 11*(11), e0165144. doi:10.1371/journal.pone.0165144
247. Roy, D. S., Arons, A., Mitchell, T. I., Pignatelli, M., Ryan, T. J., & Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer's disease. *Nature, 531*(7595), 508-512. doi:10.1038/nature17172
248. Song, N., Zhang, L., Chen, W., Zhu, H., Deng, W., Han, Y., . . . Qin, C. (2016). Cyanidin 3-O-beta-glucopyranoside activates peroxisome proliferator-activated receptor-gamma and alleviates cognitive impairment in the APP(swe)/PS1(DeltaE9) mouse model. *Biochim Biophys Acta, 1862*(9), 1786-1800. doi:10.1016/j.bbadis.2016.05.016
249. Tavares, E., Antequera, D., Lopez-Gonzalez, I., Ferrer, I., Minano, F. J., & Carro, E. (2016). Potential Role of Aminoprocalcitonin in the Pathogenesis of Alzheimer Disease. *Am J Pathol, 186*(10), 2723-2735. doi:10.1016/j.ajpath.2016.06.006
250. Wang, Q. H., Wang, Y. R., Zhang, T., Jiao, S. S., Liu, Y. H., Zeng, F., . . . Wang, Y. J. (2016). Intramuscular delivery of p75NTR ectodomain by an AAV vector attenuates cognitive deficits and Alzheimer's disease-like pathologies in APP/PS1 transgenic mice. *J Neurochem, 138*(1), 163-173. doi:10.1111/jnc.13616
251. Ahuja, M., Buabeid, M., Abdel-Rahman, E., Majrashi, M., Parameshwaran, K., Amin, R., . . . Dhanasekaran, M. (2017). Immunological alteration & toxic molecular inductions leading to cognitive impairment & neurotoxicity in transgenic mouse model of Alzheimer's disease. *Life Sci, 177*, 49-59. doi:10.1016/j.lfs.2017.03.004
252. Choi, J., & Jeong, Y. (2017). Elevated emotional contagion in a mouse model of Alzheimer's disease is associated with increased synchronization in the insula and amygdala. *Sci Rep, 7*, 46262. doi:10.1038/srep46262
253. Jin, W. S., Shen, L. L., Bu, X. L., Zhang, W. W., Chen, S. H., Huang, Z. L., . . . Wang, Y. J. (2017). Peritoneal dialysis reduces amyloid-beta plasma levels in humans and attenuates Alzheimer-associated phenotypes in an APP/PS1 mouse model. *Acta Neuropathol, 134*(2), 207-220. doi:10.1007/s00401-017-1721-y
254. Ofengeim, D., Mazzitelli, S., Ito, Y., DeWitt, J. P., Mifflin, L., Zou, C., . . . Yuan, J. (2017). RIPK1 mediates a disease-associated microglial response in Alzheimer's disease. *Proc Natl Acad Sci U S A, 114*(41), E8788-E8797. doi:10.1073/pnas.1714175114
255. van Groen, T., Schemmert, S., Brener, O., Gremer, L., Ziehm, T., Tusche, M., . . . Willbold, D. (2017). The Abeta oligomer eliminating D-enantiomeric peptide RD2 improves cognition without changing plaque pathology. *Sci Rep, 7*(1), 16275. doi:10.1038/s41598-017-16565-1
256. Vicens, P., Heredia, L., Torrente, M., & Domingo, J. L. (2017). Behavioural effects of PNU-282987 and stress in an animal model of Alzheimer's disease. *Psychogeriatrics, 17*(1), 33-42. doi:10.1111/psyg.12189
257. Yang, J., Kou, J., Lalonde, R., & Fukuchi, K. I. (2017). Intracranial IL-17A overexpression decreases cerebral amyloid angiopathy by upregulation of ABCA1 in an animal model of Alzheimer's disease. *Brain Behav Immun, 65*, 262-273. doi:10.1016/j.bbi.2017.05.012
258. Yu, L., Liu, Y., Yang, H., Zhu, X., Cao, X., Gao, J., . . . Xu, Y. (2017). PSD-93 Attenuates Amyloid-beta-Mediated Cognitive Dysfunction by Promoting the Catabolism of Amyloid-beta. *J Alzheimers Dis, 59*(3), 913-927. doi:10.3233/JAD-170320
259. Zhang, C., & Rissman, R. A. (2017). Corticotropin-releasing factor receptor-1 modulates biomarkers of DNA oxidation in Alzheimer's disease mice. *PLoS One, 12*(7), e0181367. doi:10.1371/journal.pone.0181367

260. Martinez Hernandez, A., Urbanke, H., Gillman, A. L., Lee, J., Ryazanov, S., Agbemenyah, H. Y., . . . Fischer, A. (2018). The diphenylpyrazole compound anle138b blocks Abeta channels and rescues disease phenotypes in a mouse model for amyloid pathology. *EMBO Mol Med, 10*(1), 32-47. doi:10.15252/emmm.201707825
261. Azkona, G., Levannon, D., Groner, Y., & Dierssen, M. (2010). In vivo effects of APP are not exacerbated by BACE2 co-overexpression: behavioural characterization of a double transgenic mouse model. *Amino Acids, 39*(5), 1571-1580. doi:10.1007/s00726-010-0662-8
262. Cheng, I. H., Scearce-Levie, K., Legleiter, J., Palop, J. J., Gerstein, H., Bien-Ly, N., . . . Mucke, L. (2007). Accelerating amyloid-beta fibrillization reduces oligomer levels and functional deficits in Alzheimer disease mouse models. *J Biol Chem, 282*(33), 23818-23828. doi:10.1074/jbc.M701078200
263. Meilandt, W. J., Yu, G. Q., Chin, J., Roberson, E. D., Palop, J. J., Wu, T., . . . Mucke, L. (2008). Enkephalin elevations contribute to neuronal and behavioral impairments in a transgenic mouse model of Alzheimer's disease. *J Neurosci, 28*(19), 5007-5017. doi:10.1523/JNEUROSCI.0590-08.2008
264. Meilandt, W. J., Cisse, M., Ho, K., Wu, T., Esposito, L. A., Scearce-Levie, K., . . . Mucke, L. (2009). Neprilysin overexpression inhibits plaque formation but fails to reduce pathogenic Abeta oligomers and associated cognitive deficits in human amyloid precursor protein transgenic mice. *J Neurosci, 29*(7), 1977-1986. doi:10.1523/JNEUROSCI.2984-08.2009
265. Thanopoulou, K., Fragkouli, A., Stylianopoulou, F., & Georgopoulos, S. (2010). Scavenger receptor class B type I (SR-BI) regulates perivascular macrophages and modifies amyloid pathology in an Alzheimer mouse model. *Proc Natl Acad Sci U S A, 107*(48), 20816-20821. doi:10.1073/pnas.1005888107
266. Cisse, M., Sanchez, P. E., Kim, D. H., Ho, K., Yu, G. Q., & Mucke, L. (2011). Ablation of cellular prion protein does not ameliorate abnormal neural network activity or cognitive dysfunction in the J20 line of human amyloid precursor protein transgenic mice. *J Neurosci, 31*(29), 10427-10431. doi:10.1523/JNEUROSCI.1459-11.2011
267. Sanchez, P. E., Zhu, L., Verret, L., Vossel, K. A., Orr, A. G., Cirrito, J. R., . . . Mucke, L. (2012). Levetiracetam suppresses neuronal network dysfunction and reverses synaptic and cognitive deficits in an Alzheimer's disease model. *Proc Natl Acad Sci U S A, 109*(42), E2895-2903. doi:10.1073/pnas.1121081109
268. Verret, L., Mann, E. O., Hang, G. B., Barth, A. M., Cobos, I., Ho, K., . . . Palop, J. J. (2012). Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell, 149*(3), 708-721. doi:10.1016/j.cell.2012.02.046
269. Wright, A. L., Zinn, R., Hohensinn, B., Konen, L. M., Beynon, S. B., Tan, R. P., . . . Vissel, B. (2013). Neuroinflammation and neuronal loss precede Abeta plaque deposition in the hAPP-J20 mouse model of Alzheimer's disease. *PLoS One, 8*(4), e59586. doi:10.1371/journal.pone.0059586
270. Dubal, D. B., Zhu, L., Sanchez, P. E., Worden, K., Broestl, L., Johnson, E., . . . Mucke, L. (2015). Life extension factor klotho prevents mortality and enhances cognition in hAPP transgenic mice. *J Neurosci, 35*(6), 2358-2371. doi:10.1523/JNEUROSCI.5791-12.2015
271. Hall, A. M., Throesch, B. T., Buckingham, S. C., Markwardt, S. J., Peng, Y., Wang, Q., . . . Roberson, E. D. (2015). Tau-dependent Kv4.2 depletion and dendritic hyperexcitability in a mouse model of Alzheimer's disease. *J Neurosci, 35*(15), 6221-6230. doi:10.1523/JNEUROSCI.2552-14.2015
272. Mably, A. J., Kanmert, D., Mc Donald, J. M., Liu, W., Caldarone, B. J., Lemere, C. A., . . . Walsh, D. M. (2015). Tau immunization: a cautionary tale? *Neurobiol Aging, 36*(3), 1316-1332. doi:10.1016/j.neurobiolaging.2014.11.022
273. Mably, A. J., Liu, W., Mc Donald, J. M., Dodart, J. C., Bard, F., Lemere, C. A., . . . Walsh, D. M. (2015). Anti-Abeta antibodies incapable of reducing cerebral Abeta oligomers fail to attenuate spatial reference memory deficits in J20 mice. *Neurobiol Dis, 82*, 372-384. doi:10.1016/j.nbd.2015.07.008

274. Liu, Y. L., Chen, W. T., Lin, Y. Y., Lu, P. H., Hsieh, S. L., & Cheng, I. H. (2017). Amelioration of amyloid-beta-induced deficits by DcR3 in an Alzheimer's disease model. *Mol Neurodegener, 12*(1), 30. doi:10.1186/s13024-017-0173-0
275. Orr, A. G., Lo, I., Schumacher, H., Ho, K., Gill, M., Guo, W., . . . Mucke, L. (2018). Istradefylline reduces memory deficits in aging mice with amyloid pathology. *Neurobiol Dis, 110*, 29-36. doi:10.1016/j.nbd.2017.10.014
276. Tapia-Rojas, C., & Inestrosa, N. C. (2018). Wnt signaling loss accelerates the appearance of neuropathological hallmarks of Alzheimer's disease in J20-APP transgenic and wild-type mice. *J Neurochem, 144*(4), 443-465. doi:10.1111/jnc.14278
277. Ye, X., Feng, T., Tammineni, P., Chang, Q., Jeong, Y. Y., Margolis, D. J., . . . Cai, Q. (2017). Regulation of Synaptic Amyloid-beta Generation through BACE1 Retrograde Transport in a Mouse Model of Alzheimer's Disease. *J Neurosci, 37*(10), 2639-2655. doi:10.1523/JNEUROSCI.2851-16.2017
278. Kobayashi, D., Zeller, M., Cole, T., Buttini, M., McConlogue, L., Sinha, S., . . . Chen, K. S. (2008). BACE1 gene deletion: impact on behavioral function in a model of Alzheimer's disease. *Neurobiol Aging, 29*(6), 861-873. doi:10.1016/j.neurobiolaging.2007.01.002
279. Hurtado, D. E., Molina-Porcel, L., Carroll, J. C., Macdonald, C., Aboagye, A. K., Trojanowski, J. Q., & Lee, V. M. (2012). Selectively silencing GSK-3 isoforms reduces plaques and tangles in mouse models of Alzheimer's disease. *J Neurosci, 32*(21), 7392-7402. doi:10.1523/JNEUROSCI.0889-12.2012
280. Todd Roach, J., Volmar, C. H., Dwivedi, S., Town, T., Crescentini, R., Crawford, F., . . . Mullan, M. (2004). Behavioral effects of CD40-CD40L pathway disruption in aged PSAPP mice. *Brain Res, 1015*(1-2), 161-168. doi:10.1016/j.brainres.2004.05.004
281. Jensen, M. T., Mottin, M. D., Cracchiolo, J. R., Leighty, R. E., & Arendash, G. W. (2005). Lifelong immunization with human beta-amyloid (1-42) protects Alzheimer's transgenic mice against cognitive impairment throughout aging. *Neuroscience, 130*(3), 667-684. doi:10.1016/j.neuroscience.2004.09.055
282. V., H. L. A. D. M. H. A. R. Y. K. A. R. M. M. B. (2006). Bacopa monniera extract reduces amyloid levels in PSAPP mice. *Journal of Alzheimers Disease, 9*(3), 243-251. doi:NA
283. Paris, D., Bachmeier, C., Patel, N., Quadros, A., Volmar, C. H., Laporte, V., . . . Mullan, M. J. (2011). Selective antihypertensive dihydropyridines lower Abeta accumulation by targeting both the production and the clearance of Abeta across the blood-brain barrier. *Mol Med, 17*(3-4), 149-162. doi:10.2119/molmed.2010.00180
284. Herran, E., Perez-Gonzalez, R., Igartua, M., Pedraz, J. L., Carro, E., & Hernandez, R. M. (2013). VEGF-releasing biodegradable nanospheres administered by craniotomy: a novel therapeutic approach in the APP/Ps1 mouse model of Alzheimer's disease. *J Control Release, 170*(1), 111-119. doi:10.1016/j.jconrel.2013.04.028
285. Perez-Gonzalez, R., Pascual, C., Antequera, D., Bolos, M., Redondo, M., Perez, D. I., . . . Carro, E. (2013). Phosphodiesterase 7 inhibitor reduced cognitive impairment and pathological hallmarks in a mouse model of Alzheimer's disease. *Neurobiol Aging, 34*(9), 2133-2145. doi:10.1016/j.neurobiolaging.2013.03.011
286. Le Cudennec, C., Faure, A., Ly, M., & Delatour, B. (2008). One-year longitudinal evaluation of sensorimotor functions in APP751SL transgenic mice. *Genes Brain Behav, 7 Suppl 1*, 83-91. doi:10.1111/j.1601-183X.2007.00374.x
287. Havas, D., Hutter-Paier, B., Ubhi, K., Rockenstein, E., Crailsheim, K., Masliah, E., & Windisch, M. (2011). A longitudinal study of behavioral deficits in an AbetaPP transgenic mouse model of Alzheimer's disease. *J Alzheimers Dis, 25*(2), 231-243. doi:10.3233/JAD-2011-101866
288. Faizi, M., Bader, P. L., Saw, N., Nguyen, T. V., Beraki, S., Wyss-Coray, T., . . . Shamloo, M. (2012). Thy1-hAPP(Lond/Swe+) mouse model of Alzheimer's disease displays broad behavioral deficits in sensorimotor, cognitive and social function. *Brain Behav, 2*(2), 142-154. doi:10.1002/brb3.41

289. Spencer, B., Desplats, P. A., Overk, C. R., Valera-Martin, E., Rissman, R. A., Wu, C., . . . Masliah, E. (2016). Reducing Endogenous alpha-Synuclein Mitigates the Degeneration of Selective Neuronal Populations in an Alzheimer's Disease Transgenic Mouse Model. *J Neurosci, 36*(30), 7971-7984. doi:10.1523/JNEUROSCI.0775-16.2016
290. Dhurandhar, E. J., Allison, D. B., van Groen, T., & Kadish, I. (2013). Hunger in the absence of caloric restriction improves cognition and attenuates Alzheimer's disease pathology in a mouse model. *PLoS One, 8*(4), e60437. doi:10.1371/journal.pone.0060437
291. Beitnere, U., van Groen, T., Kumar, A., Jansone, B., Klusa, V., & Kadish, I. (2014). Mildronate improves cognition and reduces amyloid-beta pathology in transgenic Alzheimer's disease mice. *J Neurosci Res, 92*(3), 338-346. doi:10.1002/jnr.23315
292. Kadish, I., Kumar, A., Beitnere, U., Jennings, E., McGilberry, W., & van Groen, T. (2016). Dietary composition affects the development of cognitive deficits in WT and Tg AD model mice. *Exp Gerontol, 86*, 39-49. doi:10.1016/j.exger.2016.05.003
293. Ziehm, T., Brener, O., van Groen, T., Kadish, I., Frenzel, D., Tusche, M., . . . Willbold, D. (2016). Increase of Positive Net Charge and Conformational Rigidity Enhances the Efficacy of d-Enantiomeric Peptides Designed to Eliminate Cytotoxic Abeta Species. *ACS Chem Neurosci, 7*(8), 1088-1096. doi:10.1021/acschemneuro.6b00047
294. Klein, A. N., Ziehm, T., van Groen, T., Kadish, I., Elfgen, A., Tusche, M., . . . Willbold, D. (2017). Optimization of d-Peptides for Abeta Monomer Binding Specificity Enhances Their Potential to Eliminate Toxic Abeta Oligomers. *ACS Chem Neurosci, 8*(9), 1889-1900. doi:10.1021/acschemneuro.7b00045
295. Scholtzova, H., Do, E., Dhakal, S., Sun, Y., Liu, S., Mehta, P. D., & Wisniewski, T. (2017). Innate Immunity Stimulation via Toll-Like Receptor 9 Ameliorates Vascular Amyloid Pathology in Tg-SwDI Mice with Associated Cognitive Benefits. *The Journal of Neuroscience, 37*(4), 936-959. doi:10.1523/jneurosci.1967-16.2017
296. Chapman, P. F., White, G. L., Jones, M. W., Cooper-Blacketer, D., Marshall, V. J., Irizarry, M., . . . Hsiao, K. K. (1999). Impaired synaptic plasticity and learning in aged amyloid precursor protein transgenic mice. *Nat Neurosci, 2*(3), 271-276. doi:10.1038/6374
297. King, D. L., Arendash, G. W., Crawford, F., Sterk, T., Menendez, J., & Mullan, M. J. (1999). Progressive and gender-dependent cognitive impairment in the APPSW transgenic mouse model for Alzheimer's disease. *Behavioural Brain Research, 103*(2), 145-162. doi:10.1016/s0166-4328(99)00037-6
298. G., L. G. P. Y. F. C. T. G. E. U. O. B. W. O. J. B. H.-A. K. F. S. A. a. C. (2001). Ibuprofen effects on Alzheimer pathology and open field activity in APPsw transgenic mice. *Neurobiol Aging, 22*(6), 983-991. doi:NA
299. X. and Nordberg A., B. I. P. D. M. A. P. T. M. S. M. H.-L. E. M. M. C. J. M. C. P. E. M. A. Z. (2002). Selective nicotinic receptor consequences in APP(SWE) transgenic mice. *Mol Cell Neurosci, 20*(2), 354-365. doi:NA
300. King, D. L., & Arendash, G. W. (2002). Behavioral characterization of the Tg2576 transgenic model of Alzheimer's disease through 19 months. *Physiology & Behavior, 75*(5), 627-642. doi:10.1016/s0031-9384(02)00639-x
301. Li, L., Cao, D., Garber, D. W., Kim, H., & Fukuchi, K.-i. (2003). Association of Aortic Atherosclerosis with Cerebral β-Amyloidosis and Learning Deficits in a Mouse Model of Alzheimer's Disease. *The American Journal of Pathology, 163*(6), 2155-2164. doi:10.1016/s0002-9440(10)63572-9
302. Li, L., Cao, D., Kim, H., Lester, R., & Fukuchi, K. (2006). Simvastatin enhances learning and memory independent of amyloid load in mice. *Ann Neurol, 60*(6), 729-739. doi:10.1002/ana.21053
303. Middei, S., Daniele, S., Caprioli, A., Ghirardi, O., & Ammassari-Teule, M. (2006). Progressive cognitive decline in a transgenic mouse model of Alzheimer's disease overexpressing mutant hAPPswe. *Genes Brain Behav, 5*(3), 249-256. doi:10.1111/j.1601-183X.2005.00160.x

304. Ribes, D., Colomina, M. T., Vicens, P., & Domingo, J. L. (2008). Effects of oral aluminum exposure on behavior and neurogenesis in a transgenic mouse model of Alzheimer's disease. *Exp Neurol, 214*(2), 293-300. doi:10.1016/j.expneurol.2008.08.017
305. Deacon, R. M., Koros, E., Bornemann, K. D., & Rawlins, J. N. (2009). Aged Tg2576 mice are impaired on social memory and open field habituation tests. *Behav Brain Res, 197*(2), 466-468. doi:10.1016/j.bbr.2008.09.042
306. Garcia, T., Ribes, D., Colomina, M. T., Cabre, M., Domingo, J. L., & Gomez, M. (2009). Evaluation of the protective role of melatonin on the behavioral effects of aluminum in a mouse model of Alzheimer's disease. *Toxicology, 265*(1-2), 49-55. doi:10.1016/j.tox.2009.09.009
307. Mitchell, J. C., Ariff, B. B., Yates, D. M., Lau, K. F., Perkinton, M. S., Rogelj, B., . . . McLoughlin, D. M. (2009). X11beta rescues memory and long-term potentiation deficits in Alzheimer's disease APPswe Tg2576 mice. *Hum Mol Genet, 18*(23), 4492-4500. doi:10.1093/hmg/ddp408
308. Quinn, J. F., Harris, C. J., Cobb, K. E., Domes, C., Ralle, M., Brewer, G., & Wadsworth, T. L. (2010). A copper-lowering strategy attenuates amyloid pathology in a transgenic mouse model of Alzheimer's disease. *J Alzheimers Dis, 21*(3), 903-914. doi:10.3233/JAD-2010-100408
309. Rustay, N. R., Cronin, E. A., Curzon, P., Markosyan, S., Bitner, R. S., Ellis, T. A., . . . Browman, K. E. (2010). Mice expressing the Swedish APP mutation on a 129 genetic background demonstrate consistent behavioral deficits and pathological markers of Alzheimer's disease. *Brain Res, 1311*, 136-147. doi:10.1016/j.brainres.2009.11.040
310. Puolivali, J., Nurmi, A., Miettinen, T. K., Soleti, A., Riccardino, F., Kalesnykas, G., . . . Paolini, M. (2011). The radical scavenger IAC (bis(1-hydroxy-2,2,6,6-tetramethyl-4-piperidinyl) decantionate) decreases mortality, enhances cognitive functions in water maze and reduces amyloid plaque burden in hAbetaPP transgenic mice. *J Alzheimers Dis, 27*(3), 499-510. doi:10.3233/JAD-2011-110881
311. Soumyanath, A., Zhong, Y. P., Henson, E., Wadsworth, T., Bishop, J., Gold, B. G., & Quinn, J. F. (2012). Centella asiatica Extract Improves Behavioral Deficits in a Mouse Model of Alzheimer's Disease: Investigation of a Possible Mechanism of Action. *Int J Alzheimers Dis, 2012*, 381974. doi:10.1155/2012/381974
312. Hanson, J. E., La, H., Plise, E., Chen, Y. H., Ding, X., Hanania, T., . . . Zhou, Q. (2013). SAHA enhances synaptic function and plasticity in vitro but has limited brain availability in vivo and does not impact cognition. *PLoS One, 8*(7), e69964. doi:10.1371/journal.pone.0069964
313. Kim, K. S., Kim, H. S., Park, J. M., Kim, H. W., Park, M. K., Lee, H. S., . . . Moon, J. (2013). Long-term immunomodulatory effect of amniotic stem cells in an Alzheimer's disease model. *Neurobiol Aging, 34*(10), 2408-2420. doi:10.1016/j.neurobiolaging.2013.03.029
314. Verret, L., Krezymon, A., Halley, H., Trouche, S., Zerwas, M., Lazouret, M., . . . Rampon, C. (2013). Transient enriched housing before amyloidosis onset sustains cognitive improvement in Tg2576 mice. *Neurobiol Aging, 34*(1), 211-225. doi:10.1016/j.neurobiolaging.2012.05.013
315. Di Paolo, C., Reverte, I., Colomina, M. T., Domingo, J. L., & Gomez, M. (2014). Chronic exposure to aluminum and melatonin through the diet: neurobehavioral effects in a transgenic mouse model of Alzheimer disease. *Food Chem Toxicol, 69*, 320-329. doi:10.1016/j.fct.2014.04.022
316. Harris, C. J., Voss, K., Murchison, C., Ralle, M., Frahler, K., Carter, R., . . . Quinn, J. F. (2014). Oral zinc reduces amyloid burden in Tg2576 mice. *J Alzheimers Dis, 41*(1), 179-192. doi:10.3233/JAD-131703
317. Liang, J., Lopez-Valdes, H. E., Martinez-Coria, H., Lindemeyer, A. K., Shen, Y., Shao, X. M., & Olsen, R. W. (2014). Dihydromyricetin ameliorates behavioral deficits and reverses neuropathology of transgenic mouse models of Alzheimer's disease. *Neurochem Res, 39*(6), 1171-1181. doi:10.1007/s11064-014-1304-4
318. Sooy, K., Noble, J., McBride, A., Binnie, M., Yau, J. L., Seckl, J. R., . . . Webster, S. P. (2015). Cognitive and Disease-Modifying Effects of 11beta-Hydroxysteroid Dehydrogenase Type 1 Inhibition in Male Tg2576 Mice, a Model of Alzheimer's Disease. *Endocrinology, 156*(12), 4592-4603. doi:10.1210/en.2015-1395

319. Subash, S., Braidy, N., Essa, M. M., Zayana, A. B., Ragini, V., Al-Adawi, S., . . . Guillemin, G. J. (2015). Long-term (15 mo) dietary supplementation with pomegranates from Oman attenuates cognitive and behavioral deficits in a transgenic mice model of Alzheimer's disease. *Nutrition, 31*(1), 223-229. doi:10.1016/j.nut.2014.06.004
320. Subash, S., Essa, M. M., Braidy, N., Awlad-Thani, K., Vaishnav, R., Al-Adawi, S., . . . Guillemin, G. J. (2015). Diet rich in date palm fruits improves memory, learning and reduces beta amyloid in transgenic mouse model of Alzheimer's disease. *J Ayurveda Integr Med, 6*(2), 111-120. doi:10.4103/0975-9476.159073
321. Subash, S., Essa, M. M., Braidy, N., Al-Jabri, A., Vaishnav, R., Al-Adawi, S., . . . Guillemin, G. J. (2016). Consumption of fig fruits grown in Oman can improve memory, anxiety, and learning skills in a transgenic mice model of Alzheimer's disease. *Nutr Neurosci, 19*(10), 475-483. doi:10.1179/1476830514Y.0000000131
322. Yang, X., Yao, C., Tian, T., Li, X., Yan, H., Wu, J., . . . Lu, Y. (2018). A novel mechanism of memory loss in Alzheimer's disease mice via the degeneration of entorhinal-CA1 synapses. *Mol Psychiatry, 23*(2), 199-210. doi:10.1038/mp.2016.151
323. Rodriguez Cruz, Y., Strehaiano, M., Rodriguez Obaya, T., Garcia Rodriguez, J. C., & Maurice, T. (2017). An Intranasal Formulation of Erythropoietin (Neuro-EPO) Prevents Memory Deficits and Amyloid Toxicity in the APPSwe Transgenic Mouse Model of Alzheimer's Disease. *J Alzheimers Dis, 55*(1), 231-248. doi:10.3233/JAD-160500
324. Touma, C., Ambree, O., Gortz, N., Keyvani, K., Lewejohann, L., Palme, R., . . . Sachser, N. (2004). Age- and sex-dependent development of adrenocortical hyperactivity in a transgenic mouse model of Alzheimer's disease. *Neurobiol Aging, 25*(7), 893-904. doi:10.1016/j.neurobiolaging.2003.09.004
325. Gortz, N., Lewejohann, L., Tomm, M., Ambree, O., Keyvani, K., Paulus, W., & Sachser, N. (2008). Effects of environmental enrichment on exploration, anxiety, and memory in female TgCRND8 Alzheimer mice. *Behav Brain Res, 191*(1), 43-48. doi:10.1016/j.bbr.2008.03.006
326. Wetzel, M. K., Naska, S., Laliberte, C. L., Rymar, V. V., Fujitani, M., Biernaskie, J. A., . . . Kaplan, D. R. (2008). p73 regulates neurodegeneration and phospho-tau accumulation during aging and Alzheimer's disease. *Neuron, 59*(5), 708-721. doi:10.1016/j.neuron.2008.07.021
327. Ambree, O., Richter, H., Sachser, N., Lewejohann, L., Dere, E., de Souza Silva, M. A., . . . Schabitz, W. R. (2009). Levodopa ameliorates learning and memory deficits in a murine model of Alzheimer's disease. *Neurobiol Aging, 30*(8), 1192-1204. doi:10.1016/j.neurobiolaging.2007.11.010
328. Dumont, M., Wille, E., Calingasan, N. Y., Nathan, C., Flint Beal, M., & Lin, M. T. (2010). N-iminoethyl-L-lysine improves memory and reduces amyloid pathology in a transgenic mouse model of amyloid deposition. *Neurochem Int, 56*(2), 345-351. doi:10.1016/j.neuint.2009.11.006
329. Dumont, M., Kipiani, K., Yu, F., Wille, E., Katz, M., Calingasan, N. Y., . . . Beal, M. F. (2011). Coenzyme Q10 decreases amyloid pathology and improves behavior in a transgenic mouse model of Alzheimer's disease. *J Alzheimers Dis, 27*(1), 211-223. doi:10.3233/JAD-2011-110209
330. Sclip, A., Antoniou, X., Colombo, A., Camici, G. G., Pozzi, L., Cardinetti, D., . . . Borsello, T. (2011). c-Jun N-terminal kinase regulates soluble Abeta oligomers and cognitive impairment in AD mouse model. *J Biol Chem, 286*(51), 43871-43880. doi:10.1074/jbc.M111.297515
331. Musilli, M., Nicolia, V., Borrelli, S., Scarpa, S., & Diana, G. (2013). Behavioral effects of Rho GTPase modulation in a model of Alzheimer's disease. *Behav Brain Res, 237*, 223-229. doi:10.1016/j.bbr.2012.09.043
332. Ma, K., & McLaurin, J. (2014). alpha-Melanocyte stimulating hormone prevents GABAergic neuronal loss and improves cognitive function in Alzheimer's disease. *J Neurosci, 34*(20), 6736-6745. doi:10.1523/JNEUROSCI.5075-13.2014
333. Walker, J. M., Klakotskaia, D., Ajit, D., Weisman, G. A., Wood, W. G., Sun, G. Y., . . . Schachtman, T. R. (2015). Beneficial effects of dietary EGCG and voluntary exercise on behavior in an Alzheimer's disease mouse model. *J Alzheimers Dis, 44*(2), 561-572. doi:10.3233/JAD-140981

334. Cavanagh, C., Tse, Y. C., Nguyen, H. B., Krantic, S., Breitner, J. C., Quirion, R., & Wong, T. P. (2016). Inhibiting tumor necrosis factor-alpha before amyloidosis prevents synaptic deficits in an Alzheimer's disease model. *Neurobiol Aging, 47*, 41-49. doi:10.1016/j.neurobiolaging.2016.07.009
335. Herring, A., Munster, Y., Metzdorf, J., Bolczek, B., Krussel, S., Krieter, D., . . . Keyvani, K. (2016). Late running is not too late against Alzheimer's pathology. *Neurobiol Dis, 94*, 44-54. doi:10.1016/j.nbd.2016.06.003
336. Maliszewska-Cyna, E., Xhima, K., & Aubert, I. (2016). A Comparative Study Evaluating the Impact of Physical Exercise on Disease Progression in a Mouse Model of Alzheimer's Disease. *J Alzheimers Dis, 53*(1), 243-257. doi:10.3233/JAD-150660
337. Xia, F., Yiu, A., Stone, S. S. D., Oh, S., Lozano, A. M., Josselyn, S. A., & Frankland, P. W. (2017). Entorhinal Cortical Deep Brain Stimulation Rescues Memory Deficits in Both Young and Old Mice Genetically Engineered to Model Alzheimer's Disease. *Neuropsychopharmacology, 42*(13), 2493-2503. doi:10.1038/npp.2017.100

5. *Summary modelling and treatment characteristics of included studies*

ID	Author	Year	Sex(es)	Model(s)	Treatments(s)
101	Halagappa	2007	Mixed group	3xTg-AD	Intermittent Food Deprivation
102	Nelson	2007	Male , Female	3xTg-AD	Paroxetine
103	Pietropaolo	2008	Female, Male	3xTg-AD	Exercise
104	Gulinello	2009	Mixed group	3xTg-AD	
105	Blanchard	2010	Male	3xTg-AD	Peptide 6
106	Gimenez-Llort	2010	Female, Male	3xTg-AD	Exercise
107	Chadwick	2011	Male	3xTg-AD	Amitriptyline-Hydrochloride
108	Garcia-Mesa	2011	Male , Female	3xTg-AD	Continously Free Access to Running Wheel in Home Cage
109	Medeiros	2011	Mixed group	3xTg-AD, Tg-SwDI	
110	Niikura	2011	Male , Female	3xTg-AD	S14G-HN
111	Bories	2012	Female, Male	3xTg-AD	
112	Filali	2012	Female	3xTg-AD	
113	Garcia-Mesa	2012	Male	3xTg-AD	
114	Rothman	2012	Male	3xTg-AD	
115	Carvalho	2013	Male	3xTg-AD	
116	Chen	2013	Male	3xTg-AD	
117	George	2013	Male	3xTg-AD	DHT
118	Gimenez-Llort	2013	Female	3xTg-AD	
119	Hebda-Bauer	2013	Female, Male	3xTg-AD	
120	Liu	2013	Male	3xTg-AD	Nicotinamide
121	Ratia	2013	Male	3xTg-AD	Huprine X
122	Yamamoto	2013	Mixed group	3xTg-AD	
123	Blazquez	2014	Male , Female	3xTg-AD	Environmental Enrichment
124	Chen	2014	Female	3xTg-AD	
125	Kazim	2014	Female	3xTg-AD	P021
126	Pietropaolo	2014	Female	3xTg-AD	
127	St-Amour	2014	Female	3xTg-AD	IVIg
128	Baeta-Corral	2015	Male	3xTg-AD	

129	Canete	2015	Male , Female	3xTg-AD	Neonatal Handling
130	Romano	2015	Male	3xTg-AD	
131	Torres-Lista	2015	Male	3xTg-AD	
132	Yu	2015	Female	3xTg-AD	Pioglitazone Hydrochloride
133	Garcia-Mesa	2016	Female	3xTg-AD	Exercise
134	Liu	2016	Male	3xTg-AD	Dipotassium N-Stearoyltyrosinate (NSTK)
135	Magistri	2016	Male	3xTg-AD	(+)-JQ1
136	Bonfili	2017	Male	3xTg-AD	
137	Branca	2017	Female	3xTg-AD	Dyrk1-Inh
138	Corpas	2017	Male, Mixed group	3xTg-AD	SIRT1
139	Durairajan	2017	Mixed group	3xTg-AD	HLJDT
140	Esquerda-Canals	2017	Female	3xTg-AD	
141	Hussain	2017	Mixed group	3xTg-AD	Laquinimod
142	Nie	2017	Mixed group	3xTg-AD	Melatonin
143	Nie	2017	Female	3xTg-AD	Rg1
144	Volmar	2017	Mixed group	3xTg-AD	M344
145	Joyashiki	2011	Male	5xFAD	Sominone
146	Hillmann	2012	Female	5xFAD	Ibuprofen
147	Corbett	2013	Male	5xFAD	Sodium Phenylacetate
148	Bhattacharya	2014	Female, Male	5xFAD	
149	Flanigan	2014	Mixed group	5xFAD	
150	Schneider	2014	Male	5xFAD	
151	Zhang	2014	Female	5xFAD	Lentiviral-miR-188-3p
152	Bhattacharya	2015	Male	5xFAD	Memogain
153	Jeong	2015	Female	5xFAD	1950 MHz Electromagnetic Fields
154	Paesler	2015	NR	5xFAD	
155	Schneider	2015	Male	5xFAD	Methylphenidate
156	Woo	2015	NR	5xFAD	

157	Grinan-Ferre	2016	Female	5xFAD	
158	Nikolaeva	2016	NR	5xFAD	CA-7050×
159	Son	2016	Female	5xFAD	RF-EMF
160	Tang	2016	Male	5xFAD	Manganese Chloride (MnCl2 4h20)
161	Ardestani	2017	Male	5xFAD	Xamoterol + Xamoterol
162	Baranger	2017	Male	5xFAD	
163	Brandscheid	2017	Male	5xFAD	
164	Braun	2017	Male	5xFAD	Vindeburnol
165	Nakagawa	2017	Male	5xFAD	
166	Sawmiller	2017	Female	5xFAD	Nutra II
167	Wu	2017	Male	5xFAD	
168	Yang	2017	NR	5xFAD	
169	Zhen	2017	Mixed group	5xFAD	Deep Brain Reachable Low Field Magnetic Stimulation (DMS)
170	He	2018	Female	5xFAD	
171	O'Leary	2018	Female	5xFAD	
172	Son	2018	Female	5xFAD	
173	Van Dam	2003	Male	APP23	
174	Dumont	2004	Female	APP23	
175	Lalonde	2005	Female	APP23	
176	Heneka	2006	Female	APP23	Dsp4
177	Arendash	2007	Male	APP23	
178	Hernandez	2010	Mixed group	APP23	
179	Terwel	2011	NR	APP23	TO901317
180	Katsouri	2016	Female	APP23	PPARγ-Coactivator-1α
181	Dineley	2002	Mixed group	APPPS1, Tg2576	
182	Liu	2008	Mixed group	APPPS1	
183	Jardanhazi-Kurutz	2010	Female	APPPS1	
184	Govindarajan	2011	NR	APPPS1	Sodium Butyrate
185	Vom Berg	2012	Mixed group	APPPS1	Peripheral P40-Specific Antibodies

186	Agis-Balboa	2013	Mixed group	APPPS1	
187	Ferguson	2013	Male	APPPS1	
188	Gao	2013	Mixed group	APPPS1	
189	Govindarajan	2013	Male	APPPS1	
190	Lo	2013	Male	APPPS1	
191	Kummer	2015	Female	APPPS1	GFT1803
192	Psotta	2015	Male	APPPS1	
193	Woo	2015	Mixed group	APPPS1	
194	Zhu	2015	Mixed group	APPPS1	Quetiapine
195	Zhou	2016	Male	APPPS1	Atorvastatin
196	Cifuentes	2017	Male	APPPS1	L-NAME + Hydalazine
197	Du	2017	Male	APPPS1	miR-124
198	Geng	2017	Male	APPPS1	Andrographolide Sulfonate
199	Kelly	2017	Male	APPPS1	
200	Liu	2017	Female	APPPS1	rAAV-zsGreen-ShDbn1
201	Wu	2017	Male	APPPS1	Colivelin
202	Zhang	2017	Mixed group	APPPS1	Sulforaphane
203	Wang	2010	Mixed group	APPSwe/PSEN1(A246E)	SPDM2 Peptide
204	Ke	2011	Male	APPSwe/PSEN1(A246E)	Exercise
205	Filali	2013	Male	APPSwe/PSEN1(A246E)	D-Serine
206	Zhang	2013	Male	APPSwe/PSEN1(A246E)	
207	Filali	2015	Male	APPSwe/PSEN1(A246E)	
208	Huang	2016	Male	APPSwe/PSEN1(A246E)	
209	Wei	2017	Male	APPSwe/PSEN1(A246E)	TRAM-34
210	Lalonde	2005	Mixed group	APPSwe/PSEN1dE9	
211	Frye	2008	Female	APPSwe/PSEN1dE9	
212	Liu	2008	NR	APPSwe/PSEN1dE9	
213	Frye	2009	Female	APPSwe/PSEN1dE9	Progesteron P4
214	Hartmann	2010	Male	APPSwe/PSEN1dE9	

215	Park	2010	Female	APPSwe/PSEN1dE9	
216	Bonardi	2011	Female	APPSwe/PSEN1dE9	
217	Butler	2011	NR	APPSwe/PSEN1dE9	Z-Phe-Ala-Diazomethylketone (PADK)
218	Tamayev	2011	Male	APPSwe/PSEN1dE9	
219	Lim	2012	Mixed group	APPSwe/PSEN1dE9	
220	Wang	2012	Female	APPSwe/PSEN1dE9	Naringin
221	Hammerschmidt	2013	Mixed group	APPSwe/PSEN1dE9	
222	Heneka	2013	NR	APPSwe/PSEN1dE9	
223	Jansen	2013	Male	APPSwe/PSEN1dE9	Multi-Nutrient Diet FC
224	Jansen	2013	Male	APPSwe/PSEN1dE9	
225	Lok	2013	Mixed group	APPSwe/PSEN1dE9	
226	Lok	2013	NR	APPSwe/PSEN1dE9	
227	Ramos-Rodriguez	2013	NR	APPSwe/PSEN1dE9	5,7-Dihydroxytryptamine (5,7-DHT)
228	Zhang	2013	Mixed group	APPSwe/PSEN1dE9	
229	Bernstein	2014	Mixed group	APPSwe/PSEN1dE9	
230	Cheng	2014	Female	APPSwe/PSEN1dE9	Triptolide
231	Hamilton	2014	NR	APPSwe/PSEN1dE9	
232	Hong	2014	Mixed group	APPSwe/PSEN1dE9	
233	Hsiao	2014	Male	APPSwe/PSEN1dE9	
234	Maroof	2014	Male	APPSwe/PSEN1dE9	
235	Wang	2014	Male	APPSwe/PSEN1dE9	Hesperidin
236	Zhang	2014	Mixed group	APPSwe/PSEN1dE9	Tubastatin
237	Akhter	2015	Male , Female	APPSwe/PSEN1dE9	Ozone
238	Frost	2015	Male	APPSwe/PSEN1dE9	pGlu-3 Antibody 07/1
239	Huang	2015	Male	APPSwe/PSEN1dE9	Single Housed
240	O'Neal-Moffitt	2015	Mixed group	APPSwe/PSEN1dE9	Melatonin
241	Wang	2015	Male	APPSwe/PSEN1dE9	Splenocytes from Aged Tg Mice
242	Li	2016	Female	APPSwe/PSEN1dE9	Human Neural Stem Cells (hNSC)
243	Manocha	2016	Male	APPSwe/PSEN1dE9	

244	Mao	2016	Female	APPSwe/PSEN1dE9	Insulin
245	Mazzitelli	2016	NR	APPSwe/PSEN1dE9	
246	Olesen	2016	Male	APPSwe/PSEN1dE9	Paroxetine
247	Roy	2016	Male	APPSwe/PSEN1dE9	
248	Song	2016	Mixed group	APPSwe/PSEN1dE9	Cyanidin 3-O-β-Glucopyranoside
249	Tavares	2016	Male	APPSwe/PSEN1dE9	Anti-NPCT
250	Wang	2016	Female	APPSwe/PSEN1dE9	AAV-ECD-Fc
251	Ahuja	2017	Male	APPSwe/PSEN1dE9	
252	Choi	2017	Male	APPSwe/PSEN1dE9	
253	Jin	2017	Female	APPSwe/PSEN1dE9	Peritoneal Dialysis
254	Ofengeim	2017	Male	APPSwe/PSEN1dE9	Nec-1s
255	van Groen	2017	Female	APPSwe/PSEN1dE9	D-Enantiomeric Peptide RD2
256	Vicens	2017	Male	APPSwe/PSEN1dE9	PNU-282987 + RS
257	Yang	2017	Male	APPSwe/PSEN1dE9	rAAV5-IL-17a
258	Yu	2017	Male	APPSwe/PSEN1dE9	
259	Zhang	2017	Mixed group	APPSwe/PSEN1dE9	R121919
260	Martinez Hernandez	2018	Male	APPSwe/PSEN1dE9	Anle138b-Containing Dry Food Pellets
261	Azkona	2010	Male	APPSweLon	
262	Cheng	2007	Male	J20	
263	Meilandt	2008	Male	J20	
264	Meilandt	2009	NR	J20	
265	Thanopoulou	2010	Male	J20	
266	Cisse	2011	Mixed group	J20	
267	Sanchez	2012	Mixed group	J20	Levetiracetam
268	Verret	2012	Mixed group	J20	
269	Wright	2013	Male	J20	
270	Dubal	2015	Mixed group	J20	
271	Hall	2015	Mixed group	J20	Levetiracetam

272	Mably	2015	Male	J20	5.00E+02
273	Mably	2015	Male	J20	46-4
274	Liu	2017	NR	J20	
275	Orr	2017	Mixed group	J20	Istradefylline
276	Tapia-Rojas	2017	Male	J20	XAV939
277	Ye	2017	Male	J20	AAV2/9-mCherry-Snapin
278	Kobayashi	2008	Mixed group	PDAPP	
279	Hurtado	2012	Female	PDAPP	
280	Roach	2004	Mixed group	PS/APP	Anti-CD40L
281	Jensen	2005	NR	PS/APP	AB-42 Peptide Vaccination Mix
282	Holcomb	2006	NR	PS/APP	Bacopa Monniera Extract
283	Paris	2011	Mixed group	PS/APP	Nilvadipine
284	Herran	2013	Female	PS/APP	Vascular Endothelial Growth Factor Nanospheres
285	Perez-Gonzalez	2013	Male	PS/APP	S14
286	Le Cudennec	2008	Male	TASD41	
287	Havas	2011	Mixed group, Male, Female	TASD41	
288	Faizi	2012	Male	TASD41	
289	Spencer	2016	Mixed group	TASD41	
290	Dhurandhar	2013	Male	Tg-SwDI	LY444711
291	Beitnere	2014	Female	Tg-SwDI	Mildronate
292	Kadish	2016	Male	Tg-SwDI	High Protein Diet
293	Ziehm	2016	Male	Tg-SwDI	cD3r Compound
294	Klein	2017	Female	Tg-SwDI	ANK6
295	Scholtzova	2017	Mixed group	Tg-SwDI	CpG ODN 1826
296	Chapman	1999	Male	Tg2576	
297	King	1999	Male , Female	Tg2576	
298	Lim	2001	Mixed group, Male, Female	Tg2576	Ibuprofen Diet

299	Bednar	2002	Female	Tg2576	
300	King	2002	Mixed group	Tg2576	
301	Li	2003	NR	Tg2576	Atherogenic Diet
302	Li	2006	Female	Tg2576	Simvastatin
303	Middei	2006	Male	Tg2576	
304	Ribes	2008	Male	Tg2576	Aluminum Lactate
305	Deacon	2009	Male	Tg2576	
306	Garcia	2009	Female	Tg2576	Aluminium
307	Mitchell	2009	Mixed group	Tg2576	
308	Quinn	2010	Female	Tg2576	Tetrathiomolybdate
309	Rustay	2010	Female	Tg2576	
310	Puolivali	2011	Female	Tg2576	IAC
311	Soumyanath	2012	Female	Tg2576	Centella Asiatica Extract (GKW)
312	Hanson	2013	Female	Tg2576	SAHA
313	Kim	2013	Female	Tg2576	Amniotic Mesenchymal Stem Cells
314	Verret	2013	Female	Tg2576	
315	Di Paolo	2014	Female	Tg2576	Citric Acid
316	Harris	2014	Female	Tg2576	Zinc Acetate
317	Liang	2014	Male	Tg2576	Dihydromyricetin
318	Sooy	2015	Male	Tg2576	UE2316 ([4-(2-Chlorophenyl-4-Fluoro-1-Piperidinyl][5-(1hpyrazol-4-Yl)-3-Thienyl]-Methanone)
319	Subash	2015	Male	Tg2576	Pomegranate Juice Extract (PJE)
320	Subash	2015	Female	Tg2576	4% Date Diet
321	Subash	2016	Female	Tg2576	Diet with 60 Kcal% Fat and 4% Fig Fruits Extract
322	Yang	2016	Male	Tg2576	
323	Cruz	2017	Male	Tg2576	Erythropoietin
324	Touma	2004	Mixed group	TgCRND8	
325	Gortz	2008	Female	TgCRND8	Enriched Housing
326	Wetzel	2008	Male	TgCRND8	

327	Ambree	2009	Male	TgCRND8	Levodopa
328	Dumont	2010	Female	TgCRND8	4 mMD-NIL (N-Iminoethyl-L-Lysine)
329	Dumont	2011	Male	TgCRND8	Coenzyme Q10
330	Sclip	2011	NR	TgCRND8	D-JNKI1
331	Musilli	2013	Mixed group	TgCRND8	CNF1
332	Ma	2014	Mixed group	TgCRND8	Alpha-MSH
333	Walker	2015	Mixed group	TgCRND8	EGCG + Exercise
334	Cavanagh	2016	Mixed group	TgCRND8	xPro159
335	Herring	2016	Female	TgCRND8	Running Wheel and Single Housing
336	Maliszewska-Cyna	2016	Mixed group	TgCRND8	Running
337	Xia	2017	Mixed group	TgCRND8	Bilateral Deep Brain Stimulation of Entorhinal Cortex

6. *OFT protocols reported in included studies*

ID	Arena length (cm)	Arena width (cm)	Arena height (cm)	Arena shape	Wall colour	Light intensity (lux)	Time habituated (minutes)	Duration (minutes)	Number of trials	Area
101	NR	NR	NR	NR	NR	NR	NR	10	NR	NR
102	47	47	NR	Rectangular	NR	NR	NR	10	1	2209
103	NR	NR	NR	NR	NR	NR	NR	60	NR	NR
104	40.5	40.5	40.5	NR	NR	NR	30	3	NR	1640.25
105	50	50	40	Rectangular	NR	NR	15	15	NR	2500
106	55	55	25	Rectangular	White	NR	NR	5	1	3025
107	NR	NR	NR	NR	NR	NR	NR	NR	1	NR
108	55	55	25	Rectangular	White	NR	NR	5	1	3025
109	45	45	30	Rectangular	NR	NR	NR	5	NR	2025
110	45	35	15	Rectangular	NR	NR	NR	5	1	1575
111	40	40	NR	Rectangular	NR	NR	NR	60	1	1600
112	76	76	40	Circular	Transparent	350	NR	5	1	4536.46
113	NR	NR	NR	NR	NR	NR	NR	NR	1	NR
114	40	40	NR	Rectangular	NR	NR	NR	5	NR	1600
115	30	30	NR	Rectangular	NR	NR	NR	8	1	900
116	50	50	NR	Rectangular	NR	NR	NR	15	1	2500
117	42	42	NR	Rectangular	NR	NR	NR	10	1	1764
118	55	55	25	Rectangular	White	NR	NR	5	1	3025
119	72	72	72	Rectangular	White	325	NR	5	1	5184
120	NR	NR	NR	NR	NR	NR	NR	30	NR	NR
121	50	50	35	Rectangular	NR	NR	NR	5	NR	2500
122	60	60	50	Circular	NR	200	NR	5	NR	2827.433
123	55	55	25	Rectangular	NR	NR	NR	5	1	3025

124	50	50	40	Rectangular	NR	NR	NR	15	1	2500
125	50	50	40	Rectangular	NR	60	NR	15	1	2500
126	NR	NR	NR	NR	NR	NR	NR	60	1	NR
127	80	80	NR	Rectangular	NR	NR	NR	60	NR	6400
128	50	50	NR	NR	White	NR	NR	NR	1	2500
129	55	55	25	NR	White	NR	NR	5	1	NR
130	28	28	15	Rectangular	White	NR	NR	30	NR	784
131	50	50	35	NR	White	20	NR	5	NR	2500
132	50	50	40	Rectangular	NR	NR	NR	15	NR	2500
133	NR	NR	NR	NR	NR	NR	NR	5	1	NR
134	50	50	38	Rectangular	NR	NR	NR	5	NR	2500
135	27	27	NR	Rectangular	NR	NR	NR	10	NR	729
136	43	43	30	Rectangular	NR	NR	NR	4	4	1849
137	NR	NR	NR	NR	NR	NR	NR	NR	1	NR
138	55	55	25	Rectangular	NR	NR	NR	5	NR	3025
139	25	25	NR	Rectangular	Transparent	NR	NR	5	1	625
140	42	42	50	Rectangular	White	NR	NR	15	NR	1764
141	44	44	44	Rectangular	NR	7	NR	10	1	1936
142	50	50	40	Rectangular	White	120	NR	5	1	2500
143	50	50	45	Rectangular	NR	120	5	5	NR	2500
144	27	27	23	Rectangular	NR	NR	NR	15	NR	729
145	34	41	18	Rectangular	NR	NR	NR	10	NR	1394
146	50	50	38	Rectangular	Grey	NR	NR	5	NR	2500
147	NR	NR	NR	Rectangular	NR	NR	NR	30	1	NR
148	50	50	NR	NR	NR	NR	NR	15	1	2500
149	92	92	47	Circular	NR	750	NR	5	1	6647.61
150	40	40	24	Rectangular	NR	NR	NR	5	NR	1600
151	NR	NR	NR	NR	NR	NR	NR	30	NR	NR
152	50	50	NR	NR	NR	NR	NR	15	1	2500

153	60	60	50	Rectangular	NR	NR	NR	30	1	3600
154	27.5	27.5	20	Rectangular	NR	NR	NR	20	3	756.25
155	40	40	24	Rectangular	NR	NR	NR	5	NR	1600
156	40	40	NR	Rectangular	NR	NR	NR	15	8	1600
157	50	50	25	Rectangular	White	NR	NR	5	2	2500
158	30	30	35	Rectangular	Transparent	NR	NR	NR	NR	900
159	NR	NR	NR	NR	NR	NR	NR	10	NR	NR
160	40	40	30	Rectangular	NR	NR	NR	5	NR	1600
161	76	76	40	Rectangular	NR	NR	NR	NR	1	5776
162	50	50	45	Rectangular	White	40	NR	10	1	2500
163	100	100	40	Rectangular	NR	NR	NR	30	1	10000
164	23	28	NR	Rectangular	NR	NR	NR	5	1	644
165	60	60	NR	Circular	NR	NR	NR	10	1	2827.433
166	40	40	NR	Rectangular	NR	NR	NR	10	NR	1600
167	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
168	40	40	3	Rectangular	NR	NR	NR	30	1	1600
169	41	41	NR	Rectangular	NR	NR	NR	30	NR	1681
170	36	36	NR	Rectangular	Transparent	NR	30	10	1	1296
171	72	72	36	Rectangular	White	NR	NR	10	NR	5184
172	60	60	50	Rectangular	NR	NR	NR	10	NR	3600
173	50	50	NR	Rectangular	NR	NR	1	10	NR	2500
174	57	33	26	Rectangular	White	NR	NR	5	3	1881
175	55	33	18	Rectangular	Transparent	NR	NR	0.5	NR	1815
176	44	44	30	Rectangular	NR	NR	NR	15	3	1936
177	81	81	28.5	Rectangular	NR	NR	NR	5	1	6561
178	43	43	NR	Rectangular	Transparent	NR	NR	30	1	1849
179	61	61	61	Rectangular	Dark	NR	NR	10	3	3721
180	45	45	NR	NR	NR	NR	NR	5	2	2025
181	43	43	NR	Rectangular	Transparent	NR	NR	30	1	1849

182	120	120	NR	Circular	White	NR	NR	5	1	11309.73
183	60	60	60	Rectangular	White	NR	NR	10	3	3600
184	80	80	20	Rectangular	Transparent	NR	NR	3	NR	6400
185	50	50	NR	NR	NR	30	30	5	1	2500
186	100	100	20	Rectangular	NR	NR	NR	5	1	10000
187	40.6	40.6	38.1	Rectangular	NR	NR	NR	30	3	1648.36
188	63	50	40	Rectangular	NR	NR	NR	5	1	3150
189	80	80	20	Rectangular	Transparent	NR	NR	5	1	6400
190	50	50	NR	Rectangular	NR	NR	1	10	NR	2500
191	61	61	61	Rectangular	NR	NR	NR	5	3	3721
192	50	50	35	Rectangular	NR	NR	NR	5	NR	2500
193	NR	NR	NR	NR	NR	NR	NR	15	NR	NR
194	91	91	NR	Rectangular	Grey	NR	NR	5	1	8281
195	50	50	39	Rectangular	White	NR	NR	5	NR	2500
196	50	50	40	Rectangular	Grey	5	NR	NR	1	2500
197	45	45	45	Rectangular	NR	NR	NR	5	1	2025
198	45	45	20	Rectangular	NR	NR	NR	5	1	2025
199	58	58	31	Circular	NR	NR	NR	5	NR	2642.079
200	50	50	40	Rectangular	White	NR	NR	5	NR	2500
201	55	55	30	Rectangular	NR	NR	1445	5	NR	3025
202	NR	120	50	Circular	Transparent	NR	NR	3	NR	11309.73
203	40.6	40.6	NR	Rectangular	Transparent	NR	NR	4	5	1648.36
204	30	30	30	Rectangular	White	NR	NR	5	1	900
205	76	76	40	Circular	Transparent	NR	NR	5	NR	4536.46
206	50	50	39	NR	White	NR	NR	5	NR	2500
207	40	20	40	Rectangular	NR	NR	NR	5	1	800
208	60	60	25	Rectangular	NR	NR	NR	10	NR	3600
209	40	40	40	Rectangular	NR	NR	NR	5	1	1600
210	50	50	38	Rectangular	White	NR	NR	5	3	2500

211	39	39	30	NR	NR	NR	NR	5	2	1521
212	100	100	55	Circular	White	12	NR	5	NR	7853.982
213	39	39	30	Rectangular	NR	NR	NR	5	2	1521
214	45	30	15	Rectangular	Grey	NR	NR	3	NR	1350
215	40	40	40	Rectangular	White	NR	NR	10	1	1600
216	30	35	30	NR	NR	NR	NR	30	1	1050
217	55	36	NR	NR	NR	NR	NR	NR	1	1980
218	40	40	60	Rectangular	Opaque	NR	10	10	1	1600
219	50	50	NR	Rectangular	White	NR	NR	5	3	2500
220	50	50	NR	Rectangular	NR	NR	NR	5	NR	2500
221	61	61	61	Rectangular	NR	25	NR	5	3	3721
222	NR	NR	NR	Rectangular	NR	25	NR	5	3	NR
223	50	50	40	Rectangular	White	NR	NR	30	1	2500
224	50	50	50	Rectangular	White	60	NR	30	NR	2500
225	80	80	28.5	Rectangular	NR	NR	NR	10	NR	6400
226	50	50	38	Rectangular	Brown	NR	NR	10	1	2500
227	22	44	NR	Rectangular	NR	NR	NR	30	1	968
228	50	50	NR	Rectangular	NR	NR	NR	5	NR	2500
229	NR	NR	NR	NR	NR	NR	60	60	1	NR
230	NR	NR	NR	NR	NR	NR	NR	NR	3	NR
231	20	20	NR	Rectangular	NR	NR	NR	120	NR	400
232	NR	40	60	Rectangular	White	NR	30	30	1	NR
233	40	40	30	Rectangular	Transparent	NR	NR	10	NR	1600
234	30	35	25	Rectangular	NR	NR	NR	30	NR	1050
235	50	50	NR	NR	NR	NR	NR	5	1	2500
236	50	50	NR	NR	NR	NR	NR	5	NR	2500
237	42	42	20	Rectangular	Transparent	NR	NR	4	1	1764
238	27	27	20	Rectangular	Transparent	NR	NR	60	1	729
239	60	60	25	Rectangular	Black	NR	NR	10	1	3600

240	45	43.9	30	Rectangular	NR	NR	NR	5	1	1975.5
241	50	50	50	Rectangular	NR	NR	NR	20	NR	2500
242	50	50	NR	Rectangular	NR	NR	NR	NR	NR	2500
243	NR	NR	NR	Rectangular	NR	NR	NR	5	NR	NR
244	81	81	28.5	Rectangular	NR	NR	NR	15	NR	6561
245	25	25	NR	Rectangular	NR	NR	NR	30	1	625
246	60	80	60	Rectangular	Brown	NR	NR	3	NR	4800
247	52	26	NR	Rectangular	NR	NR	NR	10	NR	1352
248	50	50	40	Rectangular	White	NR	NR	5	NR	2500
249	50	50	38	NR	NR	NR	NR	5	3	2500
250	NR	NR	NR	NR	NR	NR	NR	3	NR	NR
251	NR	NR	NR	NR	NR	NR	NR	50	2	NR
252	40	40	40	Rectangular	NR	NR	NR	30	1	1600
253	NR	NR	NR	NR	NR	NR	NR	3	NR	NR
254	27	27	20	Rectangular	Transparent	NR	NR	60	NR	729
255	42	42	20	NR	Transparent	NR	NR	4	NR	1764
256	60	60	50	Rectangular	Opaque	NR	NR	30	NR	3600
257	50	50	NR	Rectangular	White	NR	NR	5	3	2500
258	40	40	50	Rectangular	White	NR	NR	10	NR	1600
259	NR	NR	NR	NR	NR	NR	NR	5	4	NR
260	40	40	NR	Rectangular	NR	NR	NR	5	NR	1600
261	70	70	50	Rectangular	White	300	NR	NR	1	4900
262	41	41	38	NR	Transparent	NR	NR	15	1	1681
263	41	41	30	Rectangular	Transparent	NR	NR	15	NR	1681
264	41	41	30	Rectangular	Transparent	NR	15	15	NR	1681
265	40	40	40	Rectangular	NR	NR	NR	5	NR	1600
266	NR	NR	NR	NR	NR	NR	NR	NR	1	NR
267	41	41	30	Rectangular	Transparent	NR	60	5	6	1681
268	40	40	30	Rectangular	Transparent	NR	NR	5	NR	1600

269	40	40	NR	Rectangular	Transparent	NR	NR	10	NR	1600
270	41	30	NR	Rectangular	NR	NR	30	10	NR	1230
271	40	40	NR	NR	NR	NR	NR	10	NR	NR
272	27.3	27.3	20.3	Rectangular	NR	NR	NR	60	NR	745.29
273	27.3	27.3	20.3	Rectangular	NR	NR	NR	60	NR	745.29
274	24.32	24.32	NR	Rectangular	NR	NR	NR	15	NR	591.4624
275	41	41	30	Rectangular	Transparent	NR	60	5	NR	1681
276	72	72	37	Rectangular	White	NR	NR	10	NR	5184
277	62	56	28	Rectangular	Transparent	NR	NR	5	NR	3472
278	25	25	NR	Rectangular	Transparent	NR	NR	15	2	625
279	40	40	35	Rectangular	Black	95	NR	13	1	1600
280	100	100	NR	Circular	NR	NR	NR	10	1	7853.982
281	81	81	28.5	Rectangular	Black	NR	NR	5	1	6561
282	NR	NR	NR	NR	NR	NR	NR	50	1	NR
283	NR	NR	NR	NR	NR	NR	NR	30	NR	NR
284	NR	NR	NR	Rectangular	Black	NR	NR	NR	3	NR
285	50	50	38	Rectangular	NR	NR	NR	5	3	2500
286	50	50	30	Rectangular	Black	35	NR	10	NR	2500
287	NR	NR	NR	NR	NR	NR	60	5	NR	NR
288	76	76	NR	Rectangular	White	NR	120	10	NR	5776
289	25.5	25.5	NR	Rectangular	NR	NR	NR	15	1	650.25
290	42	42	20	Rectangular	NR	NR	NR	4	1	1764
291	42	42	20	Rectangular	Transparent	NR	NR	NR	1	1764
292	42	42	20	Circular	Transparent	NR	NR	4	NR	1385.442
293	42	42	20	Rectangular	NR	NR	NR	4	NR	1764
294	42	42	20	Rectangular	Transparent	NR	NR	4	NR	1764
295	70	70	NR	Circular	NR	NR	NR	15	NR	3848.451
296	43.1	43.1	NR	Rectangular	NR	NR	NR	5	1	1857.61
297	81	81	28.5	Rectangular	Black	NR	NR	5	1	6561

298	62	62	NR	Rectangular	Transparent	NR	NR	5	NR	3844
299	70	70	45	Rectangular	NR	NR	NR	5	1	4900
300	81	81	28.5	Rectangular	Black	NR	NR	5	1	6561
301	50	50	38	Rectangular	White	NR	NR	5	3	2500
302	50	50	38	Rectangular	White	NR	NR	5	3	2500
303	60	60	30	Circular	Striped	NR	NR	5	1	2827.433
304	100	100	47	Rectangular	Dark	NR	NR	15	1	10000
305	50	30	18	NR	Grey	NR	NR	NR	2	1500
306	100	100	47	Rectangular	Dark	NR	NR	15	1	10000
307	80	NR	60	Circular	White	NR	10	10	NR	NR
308	38	38	64	Rectangular	White	NR	NR	5	2	1444
309	42	42	30	Rectangular	NR	NR	NR	60	1	1764
310	27	27	20.3	Rectangular	NR	NR	NR	10	NR	729
311	38	38	64	Rectangular	White	NR	NR	5	6	1444
312	27	27	20.3	NR	NR	NR	60	30	NR	729
313	40	40	NR	Rectangular	Black	NR	NR	60	1	1600
314	62	62	40	Circular	White	NR	NR	10	NR	3019.071
315	60	60	50	Rectangular	NR	NR	NR	15	1	3600
316	38	38	64	Rectangular	White	NR	NR	5	20	1444
317	NR	NR	NR	NR	NR	NR	NR	10	1	NR
318	60	60	NR	Rectangular	NR	NR	60	5	NR	3600
319	45	45	40	Rectangular	White	70	NR	60	NR	2025
320	45	45	40	Rectangular	NR	70	30	60	NR	2025
321	45	45	40	Rectangular	White	NR	30	60	NR	2025
322	100	100	NR	Rectangular	NR	NR	NR	30	NR	10000
323	50	50	50	Rectangular	White	NR	30	10	1	2500
324	80	80	40	Rectangular	NR	75	NR	10	1	6400
325	80	80	80	Rectangular	NR	60	NR	10	1	6400
326	NR	NR	NR	NR	NR	NR	4	30	NR	NR

327	30	30	40	Rectangular	Transparent	7	NR	10	3	900
328	45	45	20	Rectangular	NR	NR	NR	5	1	2025
329	45	45	NR	Rectangular	NR	NR	NR	5	3	2025
330	NR	NR	NR	Rectangular	NR	NR	NR	5	NR	NR
331	80	80	50	Rectangular	Transparent	NR	NR	10	NR	6400
332	25	47	19	Rectangular	NR	NR	NR	10	NR	1175
333	40	40	40	Rectangular	Transparent	NR	30	10	1	1600
334	60	60	NR	Rectangular	Transparent	NR	NR	5	1	3600
335	52	52	30	NR	NR	NR	NR	10	1	2704
336	20	40	NR	Rectangular	NR	NR	NR	10	NR	800
337	46	46	20	Rectangular	Opaque	NR	NR	10	NR	2116

NR: not reported.

7. *RoB and study quality reporting by publication*

101	Halagappa	2007	NR	NR	NR	NR	NR	NR	NR
102	Nelson	2007	Reported	NR	NR	NR	NR	NR	NR
103	Pietropaolo	2008	NR	NR	Reported	NR	NR	Reported	NR
104	Gulinello	2009	NR	NR	Reported	NR	NR	NA	NA
105	Blanchard	2010	NR	NR	Reported	NR	NR	Reported	NR
106	Gimenez-Llort	2010	NR	Reported	Reported	Reported	NR	Reported	Reported
107	Chadwick	2011	NR	Reported	Reported	NR	NR	NR	NR
108	Garcia-Mesa	2011	NR	NR	Reported	NR	NR	NR	NR
109	Medeiros	2011	NR	NR	Reported	NR	NR	NA	NA
110	Niikura	2011	Reported	NR	NR	Reported	NR	NR	NR
111	Bories	2012	NR	NR	Reported	NR	NR	NA	NA
112	Filali	2012	NR	NR	Reported	NR	NR	NA	NA
113	Garcia-Mesa	2012	NR	NR	Reported	Reported	NR	NA	NA
114	Rothman	2012	NR	NR	Reported	Reported	NR	Reported	NR
115	Carvalho	2013	NR	NR	NR	Reported	NR	NA	NA
116	Chen	2013	NR	NR	Reported	Reported	NR	NR	NR
117	George	2013	NR	NR	Reported	Reported	NR	NR	NR
118	Gimenez-Llort	2013	NR	Reported	NR	Reported	NR	NA	NA
119	Hebda-Bauer	2013	NR	NR	Reported	NR	NR	NA	NA
120	Liu	2013	NR	NR	Reported	Reported	NR	NR	NR
121	Ratia	2013	NR	NR	Reported	NR	NR	Reported	NR
122	Yamamoto	2013	NR	NR	NR	NR	NR	NA	NA
123	Blazquez	2014	NR	NR	Reported	NR	NR	Reported	NR
124	Chen	2014	NR	NR	Reported	Reported	NR	NR	NR

125	Kazim	2014	NR	NR	NR	Reported	NR	NR	NR
126	Pietropaolo	2014	Reported	Reported	Reported	NR	NR	NA	NA
127	St-Amour	2014	NR	NR	Reported	Reported	NR	NR	NR
128	Baeta-Corral	2015	NR	Reported	NR	NR	NR	NA	NA
129	Canete	2015	NR	NR	NR	NR	NR	NR	NR
130	Romano	2015	NR	Reported	NR	Reported	NR	NA	NA
131	Torres-Lista	2015	NR	Reported	NR	NR	NR	NA	NA
132	Yu	2015	NR	NR	Reported	Reported	NR	NR	NR
133	Garcia-Mesa	2016	NR	NR	Reported	Reported	NR	NR	NR
134	Liu	2016	NR	NR	NR	Reported	NR	Reported	NR
135	Magistri	2016	Reported	NR	NR	Reported	NR	NR	NR
136	Bonfili	2017	NR	Reported	Reported	NR	NR	NA	NA
137	Branca	2017	NR	Reported	Reported	Reported	NR	NR	NR
138	Corpas	2017	NR	NR	NR	Reported	NR	NR	NR
139	Durairajan	2017	NR	NR	NR	Reported	NR	NR	NR
140	Esquerda-Canals	2017	NR	NR	Reported	Reported	NR	NA	NA
141	Hussain	2017	Reported	NR	Reported	Reported	Reported	NR	NR
142	Nie	2017	NR	NR	Reported	NR	NR	NR	NR
143	Nie	2017	NR	NR	Reported	NR	NR	NR	NR
144	Volmar	2017	Reported	NR	Reported	Reported	NR	NR	NR
145	Joyashiki	2011	NR	NR	Reported	Reported	NR	NR	NR
146	Hillmann	2012	NR	NR	Reported	Reported	NR	NR	NR
147	Corbett	2013	NR	NR	Reported	NR	NR	NR	NR
148	Bhattacharya	2014	NR	NR	Reported	Reported	NR	NA	NA
149	Flanigan	2014	NR	NR	Reported	Reported	NR	NA	NA
150	Schneider	2014	NR	NR	Reported	Reported	NR	NA	NA
151	Zhang	2014	NR	NR	Reported	Reported	NR	NR	NR
152	Bhattacharya	2015	NR	Reported	Reported	Reported	NR	NR	NR
153	Jeong	2015	NR	Reported	Reported	Reported	NR	NR	NR

154	Paesler	2015	NR	NR	Reported	Reported	NR	NA	NA
155	Schneider	2015	Reported	NR	Reported	Reported	NR	Reported	NR
156	Woo	2015	NR	NR	NR	Reported	NR	NA	NA
157	Grinan-Ferre	2016	NR	NR	Reported	Reported	NR	NA	NA
158	Nikolaeva	2016	NR	NR	NR	NR	NR	NR	NR
159	Son	2016	NR	NR	Reported	Reported	NR	NR	NR
160	Tang	2016	NR	Reported	Reported	Reported	NR	NR	NR
161	Ardestani	2017	Reported	NR	Reported	Reported	NR	No (stated)	NR
162	Baranger	2017	NR	NR	Reported	Reported	NR	NA	NA
163	Brandscheid	2017	NR	NR	NR	Reported	NR	NA	NA
164	Braun	2017	NR	NR	Reported	NR	NR	NR	NR
165	Nakagawa	2017	NR	Reported	Reported	Reported	NR	NA	NA
166	Sawmiller	2017	NR	No (stated)	Reported	Reported	NR	Reported	NR
167	Wu	2017	NR	NR	Reported	Reported	NR	NA	NA
168	Yang	2017	NR	NR	Reported	Reported	NR	NA	NA
169	Zhen	2017	NR	Reported	Reported	Reported	NR	Reported	Reported
170	He	2018	Reported	Reported	Reported	Reported	No (stated)	NA	NA
171	O'Leary	2018	Reported	NR	Reported	NR	NR	NA	NA
172	Son	2018	NR	Reported	Reported	Reported	NR	NA	NA
173	Van Dam	2003	NR	Reported	Reported	NR	NR	NA	NA
174	Dumont	2004	NR	NR	Reported	NR	NR	NA	NA
175	Lalonde	2005	NR	NR	NR	NR	NR	NA	NA
176	Heneka	2006	NR	Reported	Reported	NR	NR	NR	NR
177	Arendash	2007	NR	NR	NR	NR	NR	NA	NA
178	Hernandez	2010	NR	Reported	Reported	NR	NR	NA	NA
179	Terwel	2011	NR	NR	Reported	NR	NR	NR	NR
180	Katsouri	2016	NR	NR	NR	NR	NR	NR	NR
181	Dineley	2002	NR	Reported	Reported	NR	NR	NA	NA
182	Liu	2008	NR	NR	NR	NR	NR	NA	NA

183	Jardanhazi-Kurutz	2010	NR	Reported	Reported	NR	NR	NA	NA
184	Govindarajan	2011	NR	NR	Reported	Reported	NR	NR	NR
185	Vom Berg	2012	NR	NR	Reported	Reported	NR	NR	NR
186	Agis-Balboa	2013	NR	NR	Reported	Reported	NR	NA	NA
187	Ferguson	2013	Reported	NR	Reported	NR	NR	NA	NA
188	Gao	2013	NR	NR	Reported	NR	NR	NA	NA
189	Govindarajan	2013	NR	NR	Reported	Reported	NR	NA	NA
190	Lo	2013	NR	NR	Reported	NR	NR	NA	NA
191	Kummer	2015	NR	NR	Reported	Reported	NR	NR	NR
192	Psotta	2015	NR	NR	Reported	Reported	NR	NA	NA
193	Woo	2015	NR	NR	NR	Reported	NR	NA	NA
194	Zhu	2015	NR	Reported	Reported	Reported	NR	Reported	Reported
195	Zhou	2016	NR	Reported	NR	NR	NR	Reported	NR
196	Cifuentes	2017	NR	Reported	Reported	Reported	NR	NR	NR
197	Du	2017	NR	NR	Reported	Reported	NR	NR	NR
198	Geng	2017	NR	NR	Reported	Reported	NR	NR	NR
199	Kelly	2017	NR	Reported	NR	NR	NR	NA	NA
200	Liu	2017	NR	NR	Reported	Reported	NR	NR	NR
201	Wu	2017	Reported	NR	NR	NR	NR	Reported	NR
202	Zhang	2017	NR	NR	Reported	Reported	NR	NR	NR
203	Wang	2010	Reported	NR	Reported	NR	NR	NR	NR
204	Ke	2011	NR	NR	Reported	NR	NR	Reported	NR
205	Filali	2013	NR	NR	Reported	NR	NR	Reported	Reported
206	Zhang	2013	NR	NR	Reported	NR	NR	NA	NA
207	Filali	2015	NR	Reported	Reported	NR	NR	NA	NA
208	Huang	2016	NR	NR	NR	NR	NR	NA	NA
209	Wei	2017	NR	NR	Reported	Reported	NR	NR	NR
210	Lalonde	2005	NR	NR	Reported	NR	NR	NA	NA

211	Frye	2008	Reported	NR	Reported	NR	NR	NA	NA
212	Liu	2008	NR	NR	Reported	NR	NR	NA	NA
213	Frye	2009	Reported	NR	Reported	NR	NR	Reported	NR
214	Hartmann	2010	NR	NR	Reported	NR	NR	NA	NA
215	Park	2010	NR	NR	Reported	NR	NR	NA	NA
216	Bonardi	2011	NR	NR	Reported	Reported	NR	NA	NA
217	Butler	2011	NR	NR	Reported	Reported	NR	NR	NR
218	Tamayev	2011	NR	Reported	NR	Reported	NR	NA	NA
219	Lim	2012	NR	NR	Reported	NR	NR	NA	NA
220	Wang	2012	NR	NR	Reported	NR	NR	Reported	NR
221	Hammerschmidt	2013	NR	Reported	Reported	Reported	NR	NA	NA
222	Heneka	2013	NR	NR	NR	Reported	NR	NA	NA
223	Jansen	2013	NR	NR	Reported	Reported	NR	NR	NR
224	Jansen	2013	Reported	NR	Reported	Reported	NR	NA	NA
225	Lok	2013	NR	NR	NR	Reported	NR	NA	NA
226	Lok	2013	NR	Reported	NR	Reported	NR	NA	NA
227	Ramos-Rodriguez	2013	NR	NR	Reported	Reported	NR	NR	NR
228	Zhang	2013	NR	NR	Reported	NR	NR	NA	NA
229	Bernstein	2014	NR	Reported	Reported	NR	NR	NA	NA
230	Cheng	2014	NR	NR	NR	NR	NR	Reported	NR
231	Hamilton	2014	NR	Reported	Reported	Reported	NR	NA	NA
232	Hong	2014	NR	NR	Reported	Reported	NR	NA	NA
233	Hsiao	2014	NR	NR	Reported	Reported	NR	NA	NA
234	Maroof	2014	Reported	NR	Reported	Reported	NR	NA	NA
235	Wang	2014	NR	NR	Reported	Reported	NR	Reported	NR
236	Zhang	2014	NR	NR	Reported	Reported	NR	NR	NR
237	Akhter	2015	NR	NR	Reported	NR	NR	Reported	NR
238	Frost	2015	Reported	NR	Reported	Reported	NR	NR	NR

239	Huang	2015	NR	Reported	NR	Reported	NR	Reported	NR
240	O'Neal-Moffitt	2015	NR	Reported	Reported	Reported	NR	NR	NR
241	Wang	2015	NR	NR	NR	Reported	NR	NR	NR
242	Li	2016	NR	NR	Reported	Reported	NR	NR	NR
243	Manocha	2016	NR	NR	Reported	Reported	NR	NA	NA
244	Mao	2016	NR	NR	NR	Reported	NR	Reported	NR
245	Mazzitelli	2016	NR	NR	Reported	Reported	NR	NA	NA
246	Olesen	2016	Reported	NR	Reported	Reported	NR	NR	NR
247	Roy	2016	Reported	Reported	NR	Reported	No (stated)	NA	NA
248	Song	2016	NR	NR	Reported	Reported	NR	Reported	NR
249	Tavares	2016	NR	Reported	NR	Reported	NR	NR	NR
250	Wang	2016	NR	NR	Reported	Reported	NR	NR	NR
251	Ahuja	2017	NR	NR	Reported	NR	NR	NA	NA
252	Choi	2017	NR	NR	Reported	Reported	NR	NA	NA
253	Jin	2017	NR	NR	NR	Reported	NR	NR	NR
254	Ofengeim	2017	NR	NR	Reported	Reported	NR	NR	NR
255	van Groen	2017	NR	NR	Reported	Reported	NR	NR	NR
256	Vicens	2017	NR	NR	Reported	Reported	NR	Reported	NR
257	Yang	2017	NR	NR	Reported	Reported	NR	NR	NR
258	Yu	2017	NR	NR	Reported	Reported	NR	NA	NA
259	Zhang	2017	NR	NR	NR	NR	NR	NR	NR
260	Martinez Hernandez	2018	NR	NR	Reported	Reported	NR	NR	NR
261	Azkona	2010	NR	NR	Reported	NR	NR	NA	NA
262	Cheng	2007	NR	Reported	Reported	NR	NR	NA	NA
263	Meilandt	2008	NR	NR	Reported	NR	NR	NA	NA
264	Meilandt	2009	NR	NR	Reported	NR	NR	NA	NA
265	Thanopoulou	2010	NR	NR	Reported	Reported	NR	NA	NA
266	Cisse	2011	NR	NR	NR	NR	NR	NA	NA

267	Sanchez	2012	NR	Reported	Reported	Reported	NR	NR	Reported
268	Verret	2012	NR	Reported	Reported	NR	NR	NA	NA
269	Wright	2013	NR	NR	Reported	Reported	NR	NA	NA
270	Dubal	2015	NR	Reported	Reported	NR	NR	NA	NA
271	Hall	2015	NR	Reported	Reported	Reported	NR	NR	NR
272	Mably	2015	NR	Reported	Reported	Reported	NR	NR	NR
273	Mably	2015	NR	Reported	Reported	NR	NR	NR	NR
274	Liu	2017	NR	NR	Reported	Reported	NR	NA	NA
275	Orr	2017	Reported	Reported	Reported	Reported	Reported	Reported	Reported
276	Tapia-Rojas	2017	Reported	No (stated)	Reported	Reported	NR	Reported	No (stated)
277	Ye	2017	NR	Reported	NR	Reported	NR	NR	Reported
278	Kobayashi	2008	Reported	Reported	Reported	Reported	NR	NA	NA
279	Hurtado	2012	NR	Reported	Reported	Reported	NR	NA	NA
280	Roach	2004	Reported	NR	NR	NR	NR	NR	NR
281	Jensen	2005	NR	NR	Reported	NR	NR	NR	NR
282	Holcomb	2006	NR	NR	NR	NR	NR	NR	NR
283	Paris	2011	NR	NR	NR	Reported	NR	NR	NR
284	Herran	2013	NR	NR	NR	NR	NR	NR	NR
285	Perez-Gonzalez	2013	NR	NR	Reported	Reported	NR	NR	NR
286	Le Cudennec	2008	NR	NR	Reported	Reported	NR	NA	NA
287	Havas	2011	NR	NR	NR	NR	NR	NA	NA
288	Faizi	2012	NR	Reported	NR	NR	NR	NA	NA
289	Spencer	2016	NR	NR	Reported	Reported	NR	NA	NA
290	Dhurandhar	2013	NR	NR	Reported	Reported	NR	Reported	NR
291	Beitnere	2014	NR	Reported	Reported	Reported	NR	Reported	NR
292	Kadish	2016	NR	NR	Reported	Reported	Reported	NR	NR
293	Ziehm	2016	Reported	Reported	NR	Reported	NR	NR	NR
294	Klein	2017	NR	Reported	NR	Reported	NR	NR	NR
295	Scholtzova	2017	NR	NR	Reported	Reported	NR	NR	NR

296	Chapman	1999	NR	Reported	NR	NR	NR	NA	NA
297	King	1999	Reported	NR	NR	NR	NR	NA	NA
298	Lim	2001	NR	Reported	NR	NR	Reported	Reported	NR
299	Bednar	2002	NR	NR	NR	NR	NR	NA	NA
300	King	2002	Reported	NR	NR	NR	NR	NA	NA
301	Li	2003	Reported	NR	Reported	NR	NR	Reported	Reported
302	Li	2006	NR	NR	Reported	NR	NR	Reported	NR
303	Middei	2006	NR	NR	NR	NR	NR	NA	NA
304	Ribes	2008	Reported	NR	Reported	NR	NR	NR	NR
305	Deacon	2009	NR	NR	NR	NR	NR	NA	NA
306	Garcia	2009	NR	NR	NR	Reported	NR	NR	NR
307	Mitchell	2009	NR	Reported	NR	NR	NR	NA	NA
308	Quinn	2010	NR	NR	Reported	NR	NR	Reported	NR
309	Rustay	2010	Reported	NR	Reported	NR	NR	NA	NA
310	Puolivali	2011	NR	NR	Reported	Reported	NR	NR	NR
311	Soumyanath	2012	NR	NR	NR	Reported	NR	NR	NR
312	Hanson	2013	NR	NR	Reported	Reported	NR	NR	NR
313	Kim	2013	NR	NR	Reported	Reported	NR	NR	NR
314	Verret	2013	Reported	Reported	NR	NR	NR	NA	NA
315	Di Paolo	2014	NR	NR	Reported	Reported	NR	NR	NR
316	Harris	2014	NR	NR	NR	NR	NR	NR	NR
317	Liang	2014	NR	NR	Reported	NR	NR	NR	NR
318	Sooy	2015	Reported	NR	Reported	NR	NR	Reported	NR
319	Subash	2015	NR	NR	Reported	Reported	NR	NR	NR
320	Subash	2015	NR	NR	Reported	Reported	NR	NR	NR
321	Subash	2016	NR	NR	Reported	Reported	NR	No (stated)	NR
322	Yang	2016	NR	Reported	NR	Reported	Reported	NA	NA
323	Cruz	2017	NR	NR	NR	Reported	NR	NR	NR
324	Touma	2004	Reported	NR	Reported	NR	NR	NA	NA

325	Gortz	2008	NR	NR	Reported	NR	NR	NR	NR
326	Wetzel	2008	NR	NR	Reported	NR	NR	NA	NA
327	Ambree	2009	NR	NR	NR	NR	NR	NR	NR
328	Dumont	2010	NR	NR	Reported	NR	NR	NR	NR
329	Dumont	2011	NR	NR	Reported	NR	NR	NR	NR
330	Sclip	2011	NR	Reported	NR	NR	NR	Reported	NR
331	Musilli	2013	Reported	NR	Reported	NR	NR	Reported	NR
332	Ma	2014	NR	NR	NR	Reported	NR	Reported	NR
333	Walker	2015	Reported	NR	NR	Reported	NR	NR	NR
334	Cavanagh	2016	NR	NR	NR	Reported	NR	NR	NR
335	Herring	2016	Reported	Reported	Reported	Reported	NR	Reported	Reported
336	Maliszewska-Cyna	2016	NR	NR	Reported	Reported	NR	Reported	NR
337	Xia	2017	NR	NR	NR	Reported	NR	NR	NR

NR: not reported, NA: not applicable

*8. **Data analysis code***

Available in Open Science Framework project: https://osf.io/e4wbf/

Appendix C7

1. ***COVID-SOLES protocol***

https://doi.org/10.17605/OSF.IO/UVHGB

2. ***COVID-SOLES web application***

https://camarades.shinyapps.io/COVID-19-SOLES/

REFERENCES

Aalten, P., de Vugt, M. E., Lousberg, R., Korten, E., Jaspers, N., Senden, B., . . . Verhey, F. R. J. (2003). Behavioral Problems in Dementia: A Factor Analysis of the Neuropsychiatric Inventory. *Dementia and Geriatric Cognitive Disorders, 15*(2), 99-105. doi:10.1159/000067972

Abraham, P. (2000). Duplicate and salami publications. *J Postgrad Med, 46*(2), 67-69.

Agatonovic-Kustrin, S., Kettle, C., & Morton, D. W. (2018). A molecular approach in drug development for Alzheimer's disease. *Biomedicine & Pharmacotherapy, 106*, 553-565. doi:https://doi.org/10.1016/j.biopha.2018.06.147

Akl, E. A., Meerpohl, J. J., Elliott, J., Kahale, L. A., & Schünemann, H. J. (2017). Living systematic reviews: 4. Living guideline recommendations. *Journal of Clinical Epidemiology, 91*, 47-53. doi:https://doi.org/10.1016/j.jclinepi.2017.08.009

Al-Zubidy, A., Carver, J. C., Hale, D. P., & Hassler, E. E. (2017). Vision for SLR tooling infrastructure: Prioritizing value-added requirements. *Information and Software Technology, 91*, 72-81. doi:https://doi.org/10.1016/j.infsof.2017.06.007

Alexander, G. C., Emerson, S., & Kesselheim, A. S. (2021). Evaluation of Aducanumab for Alzheimer Disease: Scientific Evidence and Regulatory Review Involving Efficacy, Safety, and Futility. *JAMA, 325*(17), 1717-1718. doi:10.1001/jama.2021.3854

Alexandra, B. B., Jing, L., Gregers, W., & Malcolm, M. (2016). Understanding in vivo modelling of depression in non-human animals: a systematic review protocol. *Evidence-based Preclinical Medicine, 3*(2), e00024. doi:doi:10.1002/ebm2.24

Ananiadou, S., Rea, B., Okazaki, N., Procter, R., & Thomas, J. (2009). Supporting Systematic Reviews Using Text Mining. *Social Science Computer Review, 27*(4), 509-523. doi:10.1177/0894439309332293

Andrieu, S., Coley, N., Lovestone, S., Aisen, P. S., & Vellas, B. (2015). Prevention of sporadic Alzheimer's disease: lessons learned from clinical trials and future directions. *The Lancet Neurology, 14*(9), 926-944. doi:https://doi.org/10.1016/S1474-4422(15)00153-2

Asano, Y. (1986). Characteristics of open field behavior of Wistar and Sprague-Dawley rats. *Jikken Dobutsu, 35*(4), 505-508. doi:10.1538/expanim1978.35.4_505

Avey, M. T., Moher, D., Sullivan, K. J., Fergusson, D., Griffin, G., Grimshaw, J. M., . . . Canadian Critical Care Translational Biology, G. (2016). The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research. *PLoS ONE, 11*(11), e0166733. doi:10.1371/journal.pone.0166733

Bahor, Z., Liao, J., Macleod, M. R., Bannach-Brown, A., McCann, S. K., Wever, K. E., . . . Sena, E. (2017). Risk of bias reporting in the recent animal focal cerebral

ischaemia literature. *Clin Sci (Lond), 131*(20), 2525-2532. doi:10.1042/cs20160722

Bahor, Z., Liao, J., Currie, G., Ayder, C., Macleod, M., McCann, S. K., . . . Sena, C. (2021). Development and uptake of an online systematic review platform: the early years of the CAMARADES Systematic Review Facility (SyRF). *BMJ Open Science, 5*(1), e100103. doi:10.1136/bmjos-2020-100103

Balducci, C., Tonini, R., Zianni, E., Nazzaro, C., Fiordaliso, F., Salio, M., . . . Forloni, G. (2010). Cognitive deficits associated with alteration of synaptic metaplasticity precede plaque deposition in AβPP23 transgenic mice. *J Alzheimers Dis, 21*(4), 1367-1381. doi:10.3233/jad-2010-100675

Baltaci, S. B., Mogulkoc, R., & Baltaci, A. K. (2019). Molecular Mechanisms of Early and Late LTP. *Neurochem Res, 44*(2), 281-296. doi:10.1007/s11064-018-2695-4

Bandrowski, A. E., & Martone, M. E. (2016). RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods. *Neuron, 90*(3), 434-436. doi:10.1016/j.neuron.2016.04.030

Bannach-Brown, A. (2018). Preclinical Models of Depression. Retrieved from https://camarades.shinyapps.io/Preclinical-Models-of-Depression/

Bannach-Brown, A., Przybyła, P., Thomas, J., Rice, A. S. C., Ananiadou, S., Liao, J., & Macleod, M. R. (2019). Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. *Systematic Reviews, 8*(1), 23. doi:10.1186/s13643-019-0942-7

Bannach-Brown, A., Hair, K., Bahor, Z., Soliman, N., Macleod, M., & Liao, J. (2021). Technological advances in preclinical meta-research. *BMJ Open Science, 5*(1), e100131. doi:10.1136/bmjos-2020-100131

Bannach-Brown, A., Liao, J., Wegener, G., & Macleod, M. (2016). Understanding in vivo modelling of depression in non-human animals: a systematic review protocol. *Evidence-based Preclinical Medicine, 3*(2), e00024. doi:doi:10.1002/ebm2.24

Bashir, R., Surian, D., & Dunn, A. G. (2018). Time-to-update of systematic reviews relative to the availability of new evidence. *Systematic Reviews, 7*(1), 195. doi:10.1186/s13643-018-0856-9

Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med, 7*. doi:10.1371/journal.pmed.1000326

Bastrikova, N., Gardner, G. A., Reece, J. M., Jeromin, A., & Dudek, S. M. (2008). Synapse elimination accompanies functional plasticity in hippocampal neurons. *Proc Natl Acad Sci U S A, 105*(8), 3123-3127. doi:10.1073/pnas.0800027105

Bear, M. F., & Malenka, R. C. (1994). Synaptic plasticity: LTP and LTD. *Current Opinion in Neurobiology, 4*(3), 389-399. doi:https://doi.org/10.1016/0959-4388(94)90101-5

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews, 35*(3), 565-572. doi:https://doi.org/10.1016/j.neubiorev.2010.07.002

Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in Science. *Improving the Standard for Basic and Preclinical Research, 116*(1), 116-126. doi:10.1161/circresaha.114.303819

Bello, S., Krogsboll, L. T., Gruber, J., Zhao, Z. J., Fischer, D., & Hrobjartsson, A. (2014). Lack of blinding of outcome assessors in animal model experiments implies risk of observer bias. *J Clin Epidemiol, 67*(9), 973-983. doi:10.1016/j.jclinepi.2014.04.008

Bertram, L., Lill, C. M., & Tanzi, R. E. (2010). The genetics of Alzheimer disease: back to the future. *Neuron, 68*(2), 270-281. doi:10.1016/j.neuron.2010.10.013

Billard, J.-m. (2010). Long-Term Depression in the Hippocampal CA1 Area of Aged Rats, Revisited: Contribution of Temporal Constraints Related to Slice Preparation. *PLoS ONE, 5*(3), e9843. doi:10.1371/journal.pone.0009843

Bliss, T. V., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol, 232*(2), 331-356. doi:10.1113/jphysiol.1973.sp010273

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature, 361*(6407), 31-39. doi:10.1038/361031a0

Bond, N., & Di Giusto, E. (1977). Open-Field Behavior as a Function of Age, Sex, and Repeated Trials. *Psychological Reports, 41*(2), 571-574. doi:10.2466/pr0.1977.41.2.571

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open, 7*(2), e012545. doi:10.1136/bmjopen-2016-012545

Borchelt, D. R., Thinakaran, G., Eckman, C. B., Lee, M. K., Davenport, F., Ratovitsky, T., . . . Yager, D. (1996). Familial Alzheimer's disease–linked presenilin 1 variants elevate Aβ1–42/1–40 ratio in vitro and in vivo. *Neuron, 17*(5), 1005-1013.

Borg, M. S. a. A. (2010). {The RecordLinkage Package: Detecting Errors in Data. *The R Journal, 2*(2), 61-67. doi:10.32614/RJ-2010-017

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology, 66*(11), 2215-2222. doi:10.1002/asi.23329

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA, 104*(3), 240-243. doi:10.3163/1536-5050.104.3.014

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc, 104*(3), 240-243. doi:10.3163/1536-5050.104.3.014

Buchman, A. S., & Bennett, D. A. (2011). Loss of motor function in preclinical Alzheimer's disease. *Expert review of neurotherapeutics, 11*(5), 665-676. doi:10.1586/ern.11.57

Buée, L., Hof, P. R., Bouras, C., Delacourte, A., Perl, D. P., Morrison, J. H., & Fillit, H. M. (1994). Pathological alterations of the cerebral microvasculature in Alzheimer's disease and related dementing disorders. *Acta Neuropathol, 87*(5), 469-480. doi:10.1007/bf00294173

Buxbaum, J. D., Christensen, J. L., Ruefli, A. A., Greengard, P., & Loring, J. F. (1993). Expression of APP in brains of transgenic mice containing the entire human APP gene. *Biochem Biophys Res Commun, 197*(2), 639-645. doi:10.1006/bbrc.1993.2527

Carola, V., D'Olimpio, F., Brunamonti, E., Mangia, F., & Renzi, P. (2002). Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behav Brain Res, 134*(1), 49-57. doi:https://doi.org/10.1016/S0166-4328(01)00452-1

Chalmers, I., Hedges, L. V., & Cooper, H. (2002). A Brief History of Research Synthesis. *Evaluation & the Health Professions, 25*(1), 12-37. doi:10.1177/0163278702025001003

Chartier-Harlin, M.-C., Crawford, F., Houlden, H., Warren, A., Hughes, D., Fidani, L., . . . Hardy, J. (1991). Early-onset Alzheimer's disease caused by mutations at codon 717 of the β-amyloid precursor protein gene. *Nature, 353*(6347), 844-846.

Chaudhury, D., Wang, L. M., & Colwell, C. S. (2005). Circadian Regulation of Hippocampal Long-Term Potentiation. *Journal of biological rhythms, 20*(3), 225-236. doi:10.1177/0748730405276352

Chishti, M. A., Yang, D.-S., Janus, C., Phinney, A. L., Horne, P., Pearson, J., . . . Westaway, D. (2001). Early-onset Amyloid Deposition and Cognitive Deficits in Transgenic Mice Expressing a Double Mutant Form of Amyloid Precursor Protein 695 *. *Journal of Biological Chemistry, 276*(24), 21562-21570. doi:10.1074/jbc.M100710200

Chung, G. Y.-C. (2009). Towards identifying intervention arms in randomized controlled trials: extracting coordinating constructions. *J Biomed Inform, 42*. doi:10.1016/j.jbi.2008.12.011

Cohen, A. M., Ambert, K., & McDonagh, M. (2009). Cross-topic learning for work prioritization in systematic review creation and update. *J Am Med Inform Assoc, 16*(5), 690-704. doi:10.1197/jamia.M3162

Cohen, A. M. (2011). Performance of support-vector-machine-based classification on 15 systematic review topics evaluated with the WSS@95 measure. *J Am Med Inform Assoc, 18*(1), 104; author reply 104-105. doi:10.1136/jamia.2010.008177

Constantinescu, C. S., Farooqi, N., O'Brien, K., & Gran, B. (2011). Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS). *Br J Pharmacol, 164*(4), 1079-1106. doi:10.1111/j.1476-5381.2011.01302.x

Crawley, J. N., Belknap, J. K., Collins, A., Crabbe, J. C., Frankel, W., Henderson, N., . . . Paylor, R. (1997). Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology, 132*(2), 107-124. doi:10.1007/s002130050327

Créquit, P., Trinquart, L., Yavchitz, A., & Ravaud, P. (2016). Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. *BMC Medicine, 14*(1), 8. doi:10.1186/s12916-016-0555-0

Crossley, N. A., Sena, E., Goehler, J., Horn, J., van der Worp, B., Bath, P. M., . . . Dirnagl, U. (2008). Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. *Stroke, 39*(3), 929-934. doi:10.1161/strokeaha.107.498725

Cuello, A. C., Hall, H., & Do Carmo, S. (2019). Experimental Pharmacology in Transgenic Rodent Models of Alzheimer's Disease. *Frontiers in Pharmacology, 10*(189). doi:10.3389/fphar.2019.00189

Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther, 6*(4), 37. doi:10.1186/alzrt269

Currie, G. L., Angel-Scott, H., Colvin, L., Cramond, F., Hair, K., Khandoker, L., . . . Sena, E. S. (2018). Animal models of chemotherapy-induced peripheral neuropathy: a machine-assisted systematic review and meta-analysis A comprehensive summary of the field to inform robust experimental design. *bioRxiv*, 293480. doi:10.1101/293480

Currie, G. L., Angel-scott, H. N., Colvin, L., Cramond, F., Hair, K., Khandoker, L., . . . Lalu, M. (2019). Animal models of chemotherapy-induced peripheral neuropathy: A machine-assisted systematic review and meta-analysis. *PLOS Biology, 17*(5), e3000243. doi:10.1371/journal.pbio.3000243

Davis, J., Xu, F., Deane, R., Romanov, G., Previti, M. L., Zeigler, K., . . . Van Nostrand, W. E. (2004). Early-onset and robust cerebral microvascular accumulation of amyloid beta-protein in transgenic mice expressing low levels of a vasculotropic Dutch/Iowa mutant form of amyloid beta-protein precursor. *J Biol Chem, 279*(19), 20296-20306. doi:10.1074/jbc.M312946200

de Vries, R. B., Wever, K. E., Avey, M. T., Stephens, M. L., Sena, E. S., & Leenaars, M. (2014). The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. *Ilar j, 55*(3), 427-437. doi:10.1093/ilar/ilu043

de Vries, Y. A., Roest, A. M., de Jonge, P., Cuijpers, P., Munafò, M. R., & Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological medicine, 48*(15), 2453-2455. doi:10.1017/S0033291718001873

Denenberg, V. H. (1969). Open-field bheavior in the rat: what does it mean? *Ann N Y Acad Sci, 159*(3), 852-859. doi:10.1111/j.1749-6632.1969.tb12983.x

Derby, C. A., Katz, M. J., Lipton, R. B., & Hall, C. B. (2017). Trends in Dementia Incidence in a Birth Cohort Analysis of the Einstein Aging Study. *JAMA Neurol, 74*(11), 1345-1351. doi:10.1001/jamaneurol.2017.1964

Dickson, D. W., Crystal, H. A., Bevona, C., Honer, W., Vincent, I., & Davies, P. (1995). Correlations of synaptic and pathological markers with cognition of the elderly. *Neurobiol Aging, 16*(3), 285-298; discussion 298-304. doi:10.1016/0197-4580(95)00013-5

Do Carmo, S., & Cuello, A. C. (2013). Modeling Alzheimer's disease in transgenic rats. *Mol Neurodegener, 8*(1), 37. doi:10.1186/1750-1326-8-37

Egan, K., & Macleod, M. (2014). *Two decades testing interventions in transgenic mouse models of Alzheimer disease: Designing and interpreting studies for clinical trial success* (Vol. 4).

Egan, K. J., Vesterinen, H. M., Beglopoulos, V., Sena, E. S., & Macleod, M. R. (2016). From a mouse: systematic analysis reveals limitations of experiments testing interventions in Alzheimer's disease mouse models. *Evidence-based Preclinical Medicine, 3*(1), e00015. doi:10.1002/ebm2.15

Ekert, J. O., Gould, R. L., Reynolds, G., & Howard, R. J. (2018). TNF alpha inhibitors in Alzheimer's disease: A systematic review. *Int J Geriatr Psychiatry, 33*(5), 688-694. doi:10.1002/gps.4871

Elliott, J., Turner, T., Clavisi, O., Thomas, J., Higgins, J., & Mavergames, C. (2014). Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. *PLoS Med, 11*. doi:10.1371/journal.pmed.1001603

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., . . . Thomas, J. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of Clinical Epidemiology, 91*, 23-30. doi:https://doi.org/10.1016/j.jclinepi.2017.08.010

Evans, D. A. (1990). Estimated prevalence of Alzheimer's disease in the United States. *Milbank Q, 68*(2), 267-289.

Ewald, C. Y., & Li, C. (2012). Caenorhabditis elegans as a model organism to study APP function. *Experimental brain research, 217*(3-4), 397-411. doi:10.1007/s00221-011-2905-7

Ferretti, M. T., Iulita, M. F., Cavedo, E., Chiesa, P. A., Schumacher Dimech, A., Santuccione Chadha, A., . . . the Alzheimer Precision Medicine, I. (2018). Sex differences in Alzheimer disease — the gateway to precision medicine. *Nature Reviews Neurology, 14*(8), 457-469. doi:10.1038/s41582-018-0032-9

Fitzjohn, S. M., Morton, R. A., Kuenzi, F., Rosahl, T. W., Shearman, M., Lewis, H., . . . 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. *J Neurosci, 21*(13), 4691-4698. Retrieved from http://www.jneurosci.org/content/jneuro/21/13/4691.full.pdf

Foley, A. M., Ammar, Z. M., Lee, R. H., & Mitchell, C. S. (2015). Systematic review of the relationship between amyloid-beta levels and measures of transgenic mouse cognitive deficit in Alzheimer's disease. *J Alzheimers Dis, 44*(3), 787-795. doi:10.3233/jad-142208

Fonio, E., Benjamini, Y., & Golani, I. (2012). Short and Long Term Measures of Anxiety Exhibit Opposite Results. *PLoS ONE, 7*(10), e48414. doi:10.1371/journal.pone.0048414

Franco, R., & Cedazo-Minguez, A. (2014). Successful therapies for Alzheimer's disease: why so many in animal models and none in humans? *Frontiers in Pharmacology, 5*, 146. doi:10.3389/fphar.2014.00146

Games, D., Adams, D., Alessandrini, R., Barbour, R., Berthelette, P., Blackwell, C., . . . et al. (1995). Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. *Nature, 373*(6514), 523-527. doi:10.1038/373523a0

Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Akl, E. A., Beyene, J., . . . Schünemann, H. J. (2016). When and how to update systematic reviews: consensus and checklist. *BMJ, 354*, i3507. doi:10.1136/bmj.i3507

Gauthier, S., Feldman, H. H., Schneider, L. S., Wilcock, G. K., Frisoni, G. B., Hardlund, J. H., . . . Wischik, C. M. (2016). Efficacy and safety of tau-aggregation inhibitor therapy in patients with mild or moderate Alzheimer's disease: a randomised, controlled, double-blind, parallel-arm, phase 3 trial. *Lancet, 388*(10062), 2873-2884. doi:10.1016/s0140-6736(16)31275-2

Gengler, S., Hamilton, A., & Hölscher, C. (2010). Synaptic Plasticity in the Hippocampus of a APP/PS1 Mouse Model of Alzheimer's Disease Is Impaired in Old but Not Young Mice. *PLoS ONE, 5*(3), e9764. doi:10.1371/journal.pone.0009764

Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., . . . et al. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature, 349*(6311), 704-706. doi:10.1038/349704a0

Godsil, B. P., & Fanselow, M. S. (2004). Light stimulus change evokes an activity response in the rat. *Learn Behav, 32*(3), 299-310. doi:10.3758/bf03196029

Goldman, W. P., Baty, J. D., Buckles, V. D., Sahrmann, S., & Morris, J. C. (1999). Motor dysfunction in mildly demented AD individuals without extrapyramidal signs. *Neurology, 53*(5), 956-956. doi:10.1212/wnl.53.5.956

Gromer, D., Kiser, D. P., & Pauli, P. (2021). Thigmotaxis in a virtual human open field test. *Scientific Reports, 11*(1), 6670. doi:10.1038/s41598-021-85678-5

Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci U S A, 83*(13), 4913-4917. doi:10.1073/pnas.83.13.4913

Hair, K., Macleod, M. R., Sena, E. S., Howells, D., Bath, P., Irvine, C., . . . Collaboration, o. b. o. t. I. (2019). A randomised controlled trial of an

Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Research Integrity and Peer Review, 4*(1), 12. doi:10.1186/s41073-019-0069-3

Hair, K., Bahor, Z., Macleod, M., & Sena, E. (2020). Protocol: evaluating the performance of automated deduplication tools for systematic reviews. *Open Science Framework*. doi:10.17605/OSF.IO/W3MAK

Hair, K., & McCann, S. (2020). Protocol for a systematic review of preclinical systematic reviews. *Open Science Framework*. doi:10.17605/OSF.IO/XKR5G

Hall, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative Psychology, 18*(3), 385-403. doi:10.1037/h0071444

handQi, X., Yang, M., Ren, W., Jia, J., Wang, J., Han, G., & Fan, D. (2013). Find Duplicates among the PubMed, EMBASE, and Cochrane Library Databases in Systematic Review. *PLoS ONE, 8*(8), e71838. doi:10.1371/journal.pone.0071838

Hardy, J., & Allsop, D. (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in Pharmacological Sciences, 12*, 383-388.

Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science, 297*(5580), 353-356. doi:10.1126/science.1072994

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: the amyloid cascade hypothesis. *Science, 256*(5054), 184-185. doi:10.1126/science.1566067

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

Hassing, L. B., Dahl, A. K., Thorvaldsson, V., Berg, S., Gatz, M., Pedersen, N. L., & Johansson, B. (2009). Overweight in midlife and risk of dementia: a 40-year follow-up study. *International Journal of Obesity, 33*(8), 893-898. doi:10.1038/ijo.2009.104

Hearst, M. A. (1999). *Untangling text data mining. Proceedings of the 37th annual meeting of the Association for Computational Linguistics*. College Park, Maryland: Association for Computational Linguistics.

Hebb, D. (1949). The organization of behavior; a neuropsychological theory.

Helfer, B., Prosser, A., Samara, M. T., Geddes, J. R., Cipriani, A., Davis, J. M., . . . Leucht, S. (2015). Recent meta-analyses neglect previous systematic reviews and meta-analyses about the same topic: a systematic examination. *BMC Med, 13*, 82. doi:10.1186/s12916-015-0317-4

Hirst, J. A., Howick, J., Aronson, J. K., Roberts, N., Perera, R., Koshiaris, C., & Heneghan, C. (2014). The Need for Randomization in Animal Trials: An Overview of Systematic Reviews. *PLoS ONE, 9*(6), e98856. doi:10.1371/journal.pone.0098856

Holcomb, L., Gordon, M. N., McGowan, E., Yu, X., Benkovic, S., Jantzen, P., . . . Duff, K. (1998). Accelerated Alzheimer-type phenotype in transgenic mice carrying both mutant amyloid precursor protein and presenilin 1 transgenes. *Nat Med, 4*(1), 97-100. doi:10.1038/nm0198-097

Hong, S., Beja-Glasser, V. F., Nfonoyim, B. M., Frouin, A., Li, S., Ramakrishnan, S., . . . Stevens, B. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science (New York, N.Y.), 352*(6286), 712-716. doi:10.1126/science.aad8373

Hooijmans, C. R., Pasker-de Jong, P. C., de Vries, R. B., & Ritskes-Hoitinga, M. (2012). The effects of long-term omega-3 fatty acid supplementation on cognition and Alzheimer's pathology in animal models of Alzheimer's disease: a systematic review and meta-analysis. *J Alzheimers Dis, 28*(1), 191-209. doi:10.3233/jad-2011-111217

Hooijmans, C. R., & Ritskes-Hoitinga, M. (2013). Progress in Using Systematic Reviews of Animal Studies to Improve Translational Research. *PLOS Medicine, 10*(7), e1001482. doi:10.1371/journal.pmed.1001482

Horn, J., de Haan, R. J., Vermeulen, M., Luiten, P. G. M., & Limburg, M. (2001). Nimodipine in Animal Model Experiments of Focal Cerebral Ischemia. *A Systematic Review, 32*(10), 2433-2438. doi:10.1161/hs1001.096009

Hsia, A. Y., Masliah, E., McConlogue, L., Yu, G. Q., Tatsuno, G., Hu, K., . . . Mucke, L. (1999). Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proc Natl Acad Sci U S A, 96*(6), 3228-3233. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15924/pdf/pq003228.pdf

Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., . . . Cole, G. (1996). Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science, 274*(5284), 99-102. doi:10.1126/science.274.5284.99

Huang, Y.-Y., & Kandel, E. R. (2006). Age-related enhancement of a protein synthesis-dependent late phase of LTP induced by low frequency paired-pulse stimulation in hippocampus. *Learn Mem, 13*(3), 298-306. doi:10.1101/lm.166906

Huh, S., Baek, S.-J., Lee, K.-H., Whitcomb, D. J., Jo, J., Choi, S.-M., . . . Kim, B. C. (2016). The reemergence of long-term potentiation in aged Alzheimer's disease mouse model. *Scientific Reports, 6*, 29152. doi:10.1038/srep29152

Hupe, M. (2019). EndNote X9. *Journal of Electronic Resources in Medical Libraries, 16*(3-4), 117-119.

Huston, P., & Moher, D. (1996). Redundancy, disaggregation, and the integrity of medical research. *The Lancet, 347*(9007), 1024-1026. doi:https://doi.org/10.1016/S0140-6736(96)90153-1

Jackman, S. L., & Regehr, W. G. (2017). The Mechanisms and Functions of Synaptic Facilitation. *Neuron, 94*(3), 447-464. doi:https://doi.org/10.1016/j.neuron.2017.02.047

Jackson, J., Jambrina, E., Li, J., Marston, H., Menzies, F., Phillips, K., & Gilmour, G. (2019). Targeting the Synapse in Alzheimer's Disease. *Front Neurosci, 13*(735). doi:10.3389/fnins.2019.00735

Jankowsky, J. L., Fadale, D. J., Anderson, J., Xu, G. M., Gonzales, V., Jenkins, N. A., . . . Borchelt, D. R. (2003). Mutant presenilins specifically elevate the levels of the 42 residue β-amyloid peptide in vivo: evidence for augmentation of a 42-specific γ secretase. *Human Molecular Genetics, 13*(2), 159-170. doi:10.1093/hmg/ddh019

Jankowsky, J. L., & Zheng, H. (2017). Practical considerations for choosing a mouse model of Alzheimer's disease. *Mol Neurodegener, 12*(1), 89. doi:10.1186/s13024-017-0231-7

Jannot, A. S., Agoritsas, T., Gayet-Ageron, A., & Perneger, T. V. (2013). Citation bias favoring statistically significant studies was present in medical research. *J Clin Epidemiol, 66*(3), 296-301. doi:10.1016/j.jclinepi.2012.09.015

Jiang, Y., Lin, C., Meng, W., Yu, C., Cohen, A. M., & Smalheiser, N. R. (2014). Rule-based deduplication of article records from bibliographic databases. *Database (Oxford), 2014*, bat086. doi:10.1093/database/bat086

Jiang, Y., Lin, C., Meng, W., Yu, C., Cohen, A. M., & Smalheiser, N. R. (2014). Rule-based deduplication of article records from bibliographic databases. *Database, 2014*.

Kalueff, A. V., & Tuohimaa, P. (2004). Grooming analysis algorithm for neurobehavioural stress research. *Brain Research Protocols, 13*(3), 151-158. doi:https://doi.org/10.1016/j.brainresprot.2004.04.002

Kametani, H. (1988). Analysis of age-related changes in stress-induced grooming in the rat. Differential behavioral profile of adaptation to stress. *Ann N Y Acad Sci, 525*, 101-113. doi:10.1111/j.1749-6632.1988.tb38599.x

Karp, N. A., Mason, J., Beaudet, A. L., Benjamini, Y., Bower, L., Braun, R. E., . . . International Mouse Phenotyping, C. (2017). Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun, 8*(1), 15475. doi:10.1038/ncomms15475

Karran, E., & Hardy, J. (2014). A critique of the drug discovery and phase 3 clinical programs targeting the amyloid hypothesis for Alzheimer disease. *Annals of neurology, 76*(2), 185-205. doi:10.1002/ana.24188

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLOS Biology, 8*(6), e1000412. doi:10.1371/journal.pbio.1000412

Kim, S., Martinez, D., Cavedon, L., & Yencken, L. (2011). Automatic classification of sentences to support evidence based medicine. *BMC Bioinform, 12 Suppl 2*. doi:10.1186/1471-2105-12-s2-s5

Kitazawa, M., Medeiros, R., & M LaFerla, F. (2012). Transgenic mouse models of Alzheimer disease: developing a better model as a tool for therapeutic interventions. *Current pharmaceutical design, 18*(8), 1131-1147.

Klyubin, I., Walsh, D. M., Lemere, C. A., Cullen, W. K., Shankar, G. M., Betts, V., . . . Rowan, M. J. (2005). Amyloid beta protein immunotherapy neutralizes Abeta oligomers that disrupt synaptic plasticity in vivo. *Nat Med, 11*(5), 556-561. doi:10.1038/nm1234

Knight, P., Chellian, R., Wilson, R., Behnood-Rod, A., Panunzio, S., & Bruijnzeel, A. W. (2021). Sex differences in the elevated plus-maze test and large open field test in adult Wistar rats. *Pharmacology Biochemistry and Behavior, 204*, 173168. doi:https://doi.org/10.1016/j.pbb.2021.173168

Koffie, R. M., Hyman, B. T., & Spires-Jones, T. L. (2011). Alzheimer's disease: synapses gone cold. *Mol Neurodegener, 6*(1), 63. doi:10.1186/1750-1326-6-63

Korevaar, D. A., Hooft, L., & Ter Riet, G. (2011). Systematic reviews and meta-analyses of preclinical studies: Publication bias in laboratory animal experiments. *Laboratory Animals, 45*(4), 225-230. doi:10.1258/la.2011.010121

Kwon, Y., Lemieux, M., McTavish, J., & Wathen, N. (2015). Identifying and removing duplicate records from systematic review searches. *Journal of the Medical Library Association : JMLA, 103*(4), 184-188. doi:10.3163/1536-5050.103.4.004

Lalonde, R., Fukuchi, K., & Strazielle, C. (2012). APP transgenic mice for modelling behavioural and psychological symptoms of dementia (BPSD). *Neurosci Biobehav Rev, 36*(5), 1357-1375. doi:10.1016/j.neubiorev.2012.02.011

Lamb, B. T., Sisodia, S. S., Lawler, A. M., Slunt, H. H., Kitt, C. A., Kearns, W. G., . . . Gearhart, J. D. (1993). Introduction and expression of the 400 kilobase precursor amyloid protein gene in transgenic mice. *Nature Genetics, 5*(1), 22-30.

Lamb, B. T., Call, L. M., Slunt, H. H., Bardel, K. A., Lawler, A. M., Eckman, C. B., . . . Gearhart, J. D. (1997). Altered metabolism of familial Alzheimer's disease-linked amyloid precursor protein variants in yeast artificial chromosome transgenic mice. *Hum Mol Genet, 6*(9), 1535-1541. doi:10.1093/hmg/6.9.1535

Lannfelt, L., Bogdanovic, N., Appelgren, H., Axelman, K., Lilius, L., Hansson, G., . . . Winblad, B. (1994). Amyloid precursor protein mutation causes Alzheimer's disease in a Swedish family. *Neurosci Lett, 168*(1-2), 254-256. doi:10.1016/0304-3940(94)90463-4

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neuroscience, 11*, 5-5. doi:10.1186/1471-2202-11-5

Lazic, S. E., Clarke-Williams, C. J., & Munafò, M. R. (2018). What exactly is 'N' in cell culture and animal experiments? *PLOS Biology, 16*(4), e2005282. doi:10.1371/journal.pbio.2005282

Lecanu, L., & Papadopoulos, V. (2013). Modeling Alzheimer's disease with non-transgenic rat models. *Alzheimers Res Ther, 5*(3), 17. doi:10.1186/alzrt171

Lefebvre, C., Glanville, J., Wieland, L. S., Coles, B., & Weightman, A. L. (2013). Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev, 2*, 78. doi:10.1186/2046-4053-2-78

Li, C., & Götz, J. (2017). Tau-based therapies in neurodegeneration: opportunities and challenges. *Nature Reviews Drug Discovery, 16*(12), 863-883. doi:10.1038/nrd.2017.155

Li, S., Jin, M., Koeglsperger, T., Shepardson, N. E., Shankar, G. M., & Selkoe, D. J. (2011). Soluble Aβ oligomers inhibit long-term potentiation through a mechanism involving excessive activation of extrasynaptic NR2B-containing NMDA receptors. *J Neurosci, 31*(18), 6627-6638. doi:10.1523/jneurosci.0203-11.2011

Liao, J. (2017). AutoAnnotation. *GitHub repository*. doi:https://github.com/shihikoo/AutoAnnotation

Liao, J., Ananiadou, S., Currie, G. L., Howard, B. E., Rice, A., Sena, E. S., . . . Macleod, M. R. (2018). Automation of citation screening in pre-clinical systematic reviews. *bioRxiv*. doi:10.1101/280131

Liu, P.-P., Xie, Y., Meng, X.-Y., & Kang, J.-S. (2019). History and progress of hypotheses and clinical trials for Alzheimer's disease. *Signal Transduction and Targeted Therapy, 4*(1), 29. doi:10.1038/s41392-019-0063-8

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., . . . Mukadam, N. (2017). Dementia prevention, intervention, and care. *Lancet, 390*(10113), 2673-2734. doi:10.1016/s0140-6736(17)31363-6

Long, J. M., & Holtzman, D. M. (2019). Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell, 179*(2), 312-339. doi:10.1016/j.cell.2019.09.001

Lopez , P. F., L. . (2008). GROBID. *GitHub repository*. doi:https://github.com/kermitt2/grobid

Lorenzetti, D. L., & Ghali, W. A. (2013). Reference management software for systematic reviews and meta-analyses: an exploration of usage and usability. *BMC Medical Research Methodology, 13*(1), 141. doi:10.1186/1471-2288-13-141

Lyketsos, C. G., Lopez, O., Jones, B., Fitzpatrick, A. L., Breitner, J., & DeKosky, S. (2002). Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. *JAMA, 288*(12), 1475-1483. doi:10.1001/jama.288.12.1475

Lynch, M. A. (2004). Long-term potentiation and memory. *Physiol Rev, 84*(1), 87-136. doi:10.1152/physrev.00014.2003

Macleod, M. R., Lawson McLean, A., Kyriakopoulou, A., Serghiou, S., de Wilde, A., Sherratt, N., . . . Sena, E. S. (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biol, 13*(10), e1002273. doi:10.1371/journal.pbio.1002273

Mahase, E. (2021). Three FDA advisory panel members resign over approval of Alzheimer's drug. In: British Medical Journal Publishing Group.
Mandillo, S., Tucci, V., Hölter, S. M., Meziane, H., Banchaabouchi, M. A., Kallnik, M., . . . Wurst, W. (2008). Reliability, robustness, and reproducibility in mouse behavioral phenotyping: a cross-laboratory study. *Physiol Genomics, 34*(3), 243-255. doi:10.1152/physiolgenomics.90207.2008
Marchetti, C., & Marie, H. (2011). Hippocampal synaptic plasticity in Alzheimer's disease: what have we learned so far from transgenic models? *Rev Neurosci, 22*(4), 373-402. doi:10.1515/rns.2011.035
Mathis, C., Paul, S. M., & Crawley, J. N. (1994). Characterization of benzodiazepine-sensitive behaviors in the A/J and C57BL/6J inbred strains of mice. *Behavior Genetics, 24*(2), 171-180. doi:10.1007/BF01067821
Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. R., & Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature, 429*(6993), 761-766. doi:10.1038/nature02617
McCann, S. K. (2018). Antioxidants - focal ischaemia. *SyRF Project.* Retrieved from https://app.syrf.org.uk/projects/153e59fe-daa2-43db-8a43-fd9e01d650e3/detail
McGonigle, P., & Ruggeri, B. (2014). Animal models of human disease: challenges in enabling translation. *Biochem Pharmacol, 87*(1), 162-171. doi:10.1016/j.bcp.2013.08.006
McKinney, W. T., Jr., & Bunney, W. E., Jr. (1969). Animal model of depression. I. Review of evidence: implications for research. *Arch Gen Psychiatry, 21*(2), 240-248. doi:10.1001/archpsyc.1969.01740200112015
McLean, C. A., Cherny, R. A., Fraser, F. W., Fuller, S. J., Smith, M. J., Beyreuther, K., . . . Masters, C. L. (1999). Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann Neurol, 46*(6), 860-866. doi:10.1002/1531-8249(199912)46:6<860::aid-ana8>3.0.co;2-m
McReynolds, W. E., Weir, M. W., & DeFries, J. C. (1967). Open-field behavior in mice: Effect of test illumination. *Psychonomic Science, 9*(5), 277-278. doi:10.3758/BF03332220
Millar, R. B., & Anderson, M. J. (2004). Remedies for pseudoreplication. *Fisheries Research, 70*(2), 397-407. doi:https://doi.org/10.1016/j.fishres.2004.08.016
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *BMJ, 339*. doi:10.1136/bmj.b2535
Moyaho, A., & Valencia, J. (2002). Grooming and yawning trace adjustment to unfamiliar environments in laboratory Sprague-Dawley rats (Rattus norvegicus). *Journal of Comparative Psychology, 116*(3), 263-269. doi:10.1037/0735-7036.116.3.263
Mucke, L., Masliah, E., Yu, G.-Q., Mallory, M., Rockenstein, E. M., Tatsuno, G., . . . McConlogue, L. (2000). High-level neuronal expression of Aβ1–42 in wild-

type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *Journal of Neuroscience, 20*(11), 4050-4058.
Mueen Ahmed, K., & Al Dhubaib, B. (2011). Zotero: A bibliographic assistant to researcher. *Journal of Pharmacology and Pharmacotherapeutics, 2*(4), 303-305. doi:http://dx.doi.org/10.4103/0976-500X.85940
Mullan, M., Crawford, F., Axelman, K., Houlden, H., Lilius, L., Winblad, B., & Lannfelt, L. (1992). A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N–terminus of β–amyloid. *Nature Genetics, 1*(5), 345-347. doi:10.1038/ng0892-345
Mullane, K., & Williams, M. (2019). Preclinical Models of Alzheimer's Disease: Relevance and Translational Validity. *Current Protocols in Pharmacology, 84*(1), e57. doi:10.1002/cpph.57
Mullard, A. (2021). Landmark Alzheimer's drug approval confounds research community. *Nature, 594*(7863), 309-310. doi:10.1038/d41586-021-01546-2
Mullen, P. D., & Ramírez, G. (2006). The promise and pitfalls of systematic reviews. *Annu Rev Public Health, 27*, 81-102. doi:10.1146/annurev.publhealth.27.021405.102239
Murrell, J., Farlow, M., Ghetti, B., & Benson, M. (1991). A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. *Science, 254*(5028), 97-99. doi:10.1126/science.1925564
Myers, A., & McGonigle, P. (2019). Overview of Transgenic Mouse Models for Alzheimer's Disease. *Curr Protoc Neurosci, 89*(1), e81. doi:https://doi.org/10.1002/cpns.81
Nägerl, U. V., Eberhorn, N., Cambridge, S. B., & Bonhoeffer, T. (2004). Bidirectional activity-dependent morphological plasticity in hippocampal neurons. *Neuron, 44*(5), 759-767. doi:10.1016/j.neuron.2004.11.016
Nagy, Z. M., & Glaser, H. D. (1970). Open-field behavior of C57BL/6J mice: Effect of illumination, age, and number of test days. *Psychonomic Science, 19*(3), 143-145. doi:10.3758/BF03335518
Nama, N., Sampson, M., Barrowman, N., Sandarage, R., Menon, K., Macartney, G., . . . McNally, J. D. (2019). Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study. *J Med Internet Res, 21*(4), e12953. doi:10.2196/12953
Nestler, E. J., Hyman, S. E., & Malenka, R. C. (2001). *Molecular neuropharmacology: a foundation for clinical neuroscience*: McGraw-Hill Medical.
Nistico, R., Pignatelli, M., Piccinin, S., Mercuri, N. B., & Collingridge, G. (2012). Targeting synaptic dysfunction in Alzheimer's disease therapy. *Mol Neurobiol, 46*(3), 572-587. doi:10.1007/s12035-012-8324-3
Niu, H., Alvarez-Alvarez, I., Guillen-Grima, F., & Aguinaga-Ontoso, I. (2017). Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurologia, 32*(8), 523-532. doi:10.1016/j.nrl.2016.02.016

Noel-Storr, A., Thomas, J., Mavergames, C., Dooley, G., Wisniewski, S., Turner, T., . . . Elliott, J. (2017). *Cochrane Crowd: Using citizen science to meet the challenge of information overload in evidence production*.

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Syst Rev, 4*. doi:10.1186/2046-4053-4-5

Oakley, H., Cole, S. L., Logan, S., Maus, E., Shao, P., Craft, J., . . . Vassar, R. (2006). Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 26*(40), 10129-10140. doi:10.1523/JNEUROSCI.1202-06.2006

Oddo, S., Caccamo, A., Shepherd, J. D., Murphy, M. P., Golde, T. E., Kayed, R., . . . LaFerla, F. M. (2003). Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Aβ and synaptic dysfunction. *Neuron, 39*(3), 409-421.

Oxman, A. D., & Guyatt, G. H. (1993). The science of reviewing research. *Ann N Y Acad Sci, 703*. doi:10.1111/j.1749-6632.1993.tb26342.x

Paul, L., Michael, R., & Daniel, T. (2015). The Contributions of MEDLINE, Other Bibliographic Databases and Various Search Techniques to NICE Public Health Guidance. *Evidence Based Library and Information Practice, 10*(1). doi:10.18438/B82P55

Peers, I. S., Ceuppens, P. R., & Harbron, C. (2012). In search of preclinical robustness. *Nature Reviews Drug Discovery, 11*, 733. doi:10.1038/nrd3849

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., . . . Würbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLOS Biology, 18*(7), e3000410. doi:10.1371/journal.pbio.3000410

Pound, P., Ebrahim, S., Sandercock, P., Bracken, M. B., & Roberts, I. (2004). Where is the evidence that animal research benefits humans? *BMJ, 328*(7438), 514-517. doi:10.1136/bmj.328.7438.514

Pressler, R., & Auvin, S. (2013). Comparison of Brain Maturation among Species: An Example in Translational Research Suggesting the Possible Use of Bumetanide in Newborn. *Front Neurol, 4*, 36. doi:10.3389/fneur.2013.00036

Proitsi, P., Lupton, M. K., Velayudhan, L., Newhouse, S., Fogh, I., Tsolaki, M., . . . Powell, J. F. (2014). Genetic predisposition to increased blood cholesterol and triglyceride lipid levels and risk of Alzheimer disease: a Mendelian randomization analysis. *PLoS Med, 11*(9), e1001713. doi:10.1371/journal.pmed.1001713

Prut, L., & Belzung, C. (2003). The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *Eur J Pharmacol, 463*(1), 3-33. doi:https://doi.org/10.1016/S0014-2999(03)01272-X

Qi, X.-S., Bai, M., Yang, Z.-P., & Ren, W.-R. (2013). Duplicates in systematic reviews: A critical, but often neglected issue. *World Journal of Meta-Analysis, 1*(3), 97-101.

Quinn, J. F. (2018). Lost in Translation? Finding Our Way To Effective Alzheimer's Disease Therapies. *J Alzheimers Dis, 64*(s1), S33-s39. doi:10.3233/jad-179930

Radde, R., Bolmont, T., Kaeser, S. A., Coomaraswamy, J., Lindau, D., Stoltze, L., . . . Jucker, M. (2006). Abeta42-driven cerebral amyloidosis in transgenic mice reveals early and robust pathology. *EMBO Rep, 7*(9), 940-946. doi:10.1038/sj.embor.7400784

Rae, E. A., & Brown, R. E. (2015). The problem of genotype and sex differences in life expectancy in transgenic AD mice. *Neuroscience & Biobehavioral Reviews, 57*, 238-251. doi:https://doi.org/10.1016/j.neubiorev.2015.09.002

Rathbone, J., Carter, M., Hoffmann, T., & Glasziou, P. (2015). Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews, 4*(1), 6. doi:10.1186/2046-4053-4-6

Reichlin, T. S., Vogt, L., & Würbel, H. (2016). The Researchers' View of Scientific Rigor-Survey on the Conduct and Reporting of In Vivo Research. *PLoS ONE, 11*(12), e0165999. doi:10.1371/journal.pone.0165999

Rice, D. P., Fillit, H. M., Max, W., Knopman, D. S., Lloyd, J. R., & Duttagupta, S. (2001). Prevalence, costs, and treatment of Alzheimer's disease and related dementia: a managed care perspective. *Am J Manag Care, 7*(8), 809-818.

Rinaldi, A. (2018). Setbacks and promises for drugs against Alzheimer's disease. *EMBO reports, 19*(9), e46714. doi:10.15252/embr.201846714

Robinson, J. L., Molina-Porcel, L., Corrada, M. M., Raible, K., Lee, E. B., Lee, V. M., . . . Trojanowski, J. Q. (2014). Perforant path synaptic loss correlates with cognitive impairment and Alzheimer's disease in the oldest-old. *Brain, 137*(Pt 9), 2578-2587. doi:10.1093/brain/awu190

Rockenstein, E., Mallory, M., Mante, M., Sisk, A., & Masliaha, E. (2001). Early formation of mature amyloid-beta protein deposits in a mutant APP transgenic model depends on levels of Abeta(1-42). *J Neurosci Res, 66*(4), 573-582. doi:10.1002/jnr.1247

Rodgers, S. P., Born, H. A., Das, P., & Jankowsky, J. L. (2012). Transgenic APP expression during postnatal development causes persistent locomotor hyperactivity in the adult. *Mol Neurodegener, 7*(1), 28. doi:10.1186/1750-1326-7-28

Rogers, J. T., Liu, C.-C., Zhao, N., Wang, J., Putzke, T., Yang, L., . . . Bu, G. (2017). Subacute ibuprofen treatment rescues the synaptic and cognitive deficits in advanced-aged mice. *Neurobiol Aging, 53*, 112-121. doi:https://doi.org/10.1016/j.neurobiolaging.2017.02.001

Royle, P., & Milne, R. (2003). Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care, 19*(4), 591-603.

Rubinow, M. J., Arseneau, L. M., Beverly, J. L., & Juraska, J. M. (2004). Effect of the estrous cycle on water maze acquisition depends on the temperature of the water. *Behav Neurosci, 118*(4), 863-868. doi:10.1037/0735-7044.118.4.863

Sams-Dodd, F. (2006). Strategies to optimize the validity of disease models in the drug discovery process. *Drug discovery today, 11*(7), 355-363. doi:https://doi.org/10.1016/j.drudis.2006.02.005

Sandercock, P., & Roberts, I. (2002). Systematic reviews of animal experiments. *The Lancet, 360*(9333), 586. doi:10.1016/S0140-6736(02)09812-4

Sanes, J. R., & Lichtman, J. W. (1999). Can molecules explain long-term potentiation? *Nature Neuroscience, 2*, 597. doi:10.1038/10154

Santos, L. E., Beckman, D., & Ferreira, S. T. (2016). Microglial dysfunction connects depression and Alzheimer's disease. *Brain, Behavior, and Immunity, 55*, 151-165. doi:https://doi.org/10.1016/j.bbi.2015.11.011

Sasaguri, H., Nilsson, P., Hashimoto, S., Nagata, K., Saito, T., De Strooper, B., . . . Saido, T. C. (2017). APP mouse models for Alzheimer's disease preclinical studies. *The EMBO journal, 36*(17), 2473-2487. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5579350/pdf/EMBJ-36-2473.pdf

Scheff, S. W., Price, D. A., Schmitt, F. A., & Mufson, E. J. (2006). Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging, 27*(10), 1372-1384. doi:10.1016/j.neurobiolaging.2005.09.012

Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., . . . Heywood, J. A. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. *Amyotroph Lateral Scler, 9*(1), 4-15. doi:10.1080/17482960701856300

Seblova, D., Quiroga, M. L., Fors, S., Johnell, K., Lövdén, M., de Leon, A. P., . . . Lager, A. (2018). Thirty-year trends in dementia: a nationwide population study of Swedish inpatient records. *Clin Epidemiol, 10*, 1679-1693. doi:10.2147/clep.S178955

Seibenhener, M. L., & Wooten, M. C. (2015). Use of the Open Field Maze to measure locomotor and anxiety-like behavior in mice. *Journal of Visualized Experiments : JoVE*(96), e52434-e52434. doi:10.3791/52434

Selkoe, D. J. (2002). Alzheimer's Disease Is a Synaptic Failure. *Science, 298*(5594), 789. Retrieved from http://science.sciencemag.org/content/298/5594/789.abstract http://science.sciencemag.org/content/298/5594/789.long

Sena, E. S., van der Worp, H. B., Bath, P. M. W., Howells, D. W., & Macleod, M. R. (2010). Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. *PLOS Biology, 8*(3), e1000344. doi:10.1371/journal.pbio.1000344

Sena, E. S., Currie, G. L., McCann, S. K., Macleod, M. R., & Howells, D. W. (2014). Systematic reviews and meta-analysis of preclinical studies: why perform

them and how to appraise them critically. *J Cereb Blood Flow Metab, 34*(5), 737-742. doi:10.1038/jcbfm.2014.28

Senanarong, V., Cummings, J. L., Fairbanks, L., Mega, M., Masterman, D. M., O'Connor, S. M., & Strickland, T. L. (2004). Agitation in Alzheimer's Disease Is a Manifestation of Frontal Lobe Dysfunction. *Dementia and Geriatric Cognitive Disorders, 17*(1-2), 14-20. doi:10.1159/000074080

Serra, L., Perri, R., Fadda, L., Padovani, A., Lorusso, S., Pettenati, C., . . . Carlesimo, G. A. (2010). Relationship between cognitive impairment and behavioural disturbances in Alzheimer's disease patients. *Behav Neurol, 23*(3), 123-130. doi:10.3233/ben-2010-0275

Serrano-Pozo, A., Frosch, M. P., Masliah, E., & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med, 1*(1), a006189. doi:10.1101/cshperspect.a006189

Shankar, G. M., Li, S., Mehta, T. H., Garcia-Munoz, A., Shepardson, N. E., Smith, I., . . . Selkoe, D. J. (2008). Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat Med, 14*(8), 837-842. doi:10.1038/nm1782

Shineman, D. W., Basi, G. S., Bizon, J. L., Colton, C. A., Greenberg, B. D., Hollister, B. A., . . . Fillit, H. M. (2011). Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. *Alzheimers Res Ther, 3*(5), 28. doi:10.1186/alzrt90

Shojania, K. G., Sampson, M., Ansari, M. T., Ji, J., Doucette, S., & Moher, D. (2007). How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med, 147*. doi:10.7326/0003-4819-147-4-200708210-00179

Shojania, K. G., Sampson, M., Ansari, M. T., Ji, J., Garritty, C., Rader, T., & Moher, D. (2007). AHRQ Technical Reviews. In *Updating Systematic Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US).

Shree, S., Bhardwaj, R., Kashish, & Deshmukh, R. (2017). Non-transgenic Animal Models of Alzheimer's Disease. In P. K. Bansal & R. Deshmukh (Eds.), *Animal Models of Neurological Disorders: Principle and Working Procedure for Animal Models of Neurological Disorders* (pp. 3-22). Singapore: Springer Singapore.

Simonato, M., Iyengar, S., Brooks-Kayal, A., Collins, S., Depaulis, A., Howells, D. W., . . . Sena, E. S. (2017). Identification and characterization of outcome measures reported in animal models of epilepsy: Protocol for a systematic review of the literature-A TASK2 report of the AES/ILAE Translational Task Force of the ILAE. *Epilepsia, 58 Suppl 4*, 68-77. doi:10.1111/epi.13908

Siontis, K. C., Hernandez-Boussard, T., & Ioannidis, J. P. (2013). Overlapping meta-analyses on the same topic: survey of published studies. *BMJ, 347*, f4501. doi:10.1136/bmj.f4501

Smalheiser, N. R., Lin, C., Jia, L., Jiang, Y., Cohen, A. M., Yu, C., . . . Meng, W. (2014). Design and implementation of Metta, a metasearch engine for biomedical

literature retrieval intended for systematic reviewers. *Health Inf Sci Syst, 2*. doi:10.1186/2047-2501-2-1

Spruijt, B. M., Peters, S. M., de Heer, R. C., Pothuizen, H. H., & van der Harst, J. E. (2014). Reproducibility and relevance of future behavioral sciences should benefit from a cross fertilization of past recommendations and today's technology: "Back to the future". *J Neurosci Methods, 234*, 2-12. doi:10.1016/j.jneumeth.2014.03.001

Stanford, S. C. (2007). The Open Field Test: reinventing the wheel. *Journal of Psychopharmacology, 21*(2), 134-135. doi:10.1177/0269881107073199

Sturchler-Pierrat, C., Abramowski, D., Duke, M., Wiederhold, K.-H., Mistl, C., Rothacher, S., . . . Paganetti, P. A. (1997). Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology. *Proceedings of the National Academy of Sciences, 94*(24), 13287-13292.

Sturman, O., Germain, P.-L., & Bohacek, J. (2018). Exploratory rearing: a context- and stress-sensitive behavior recorded in the open-field test. *Stress, 21*(5), 443-452. doi:10.1080/10253890.2018.1438405

Tanaka, S., Young, J. W., Halberstadt, A. L., Masten, V. L., & Geyer, M. A. (2012). Four factors underlying mouse behavior in an open field. *Behav Brain Res, 233*(1), 55-61. doi:https://doi.org/10.1016/j.bbr.2012.04.045

Tanriver-Ayder, E., Faes, C., van de Casteele, T., McCann, S. K., & Macleod, M. R. (2021). Comparison of commonly used methods in random effects meta-analysis: application to preclinical data in drug discovery research. *BMJ Open Science, 5*(1), e100074. doi:10.1136/bmjos-2020-100074

Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., DeTeresa, R., Hill, R., . . . Katzman, R. (1991). Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann Neurol, 30*(4), 572-580. doi:10.1002/ana.410300410

The NPQIP Collaborative Group, MacLeod, M., Sena, E., & Howells, D. (2019). Did a change in Nature journals' editorial policy for life sciences research improve reporting? *BMJ Open Science, 3*(1), e000035. doi:10.1136/bmjos-2017-000035

Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., . . . Elliott, J. (2017). Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology, 91*, 31-37. doi:https://doi.org/10.1016/j.jclinepi.2017.08.011

Tomiyama, T., Nagata, T., Shimada, H., Teraoka, R., Fukushima, A., Kanemitsu, H., . . . Ataka, S. (2008). A new amyloid β variant favoring oligomerization in Alzheimer's-type dementia. *Annals of neurology, 63*(3), 377-387.

Tramer, M. R., Reynolds, D. J., Moore, R. A., & McQuay, H. J. (1997). Impact of covert duplicate publication on meta-analysis: a case study. *BMJ, 315*(7109), 635-640. doi:10.1136/bmj.315.7109.635

Tricco, A. C., Brehaut, J., Chen, M. H., & Moher, D. (2008). Following 411 Cochrane Protocols to Completion: A Retrospective Cohort Study. *PLoS ONE, 3*(11), e3684. doi:10.1371/journal.pone.0003684

Tripathy, S. J., Burton, S. D., Geramita, M., Gerkin, R. C., & Urban, N. N. (2015). Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. *Journal of Neurophysiology, 113*(10), 3474-3489. doi:10.1152/jn.00237.2015

Trullas, R., & Skolnick, P. (1993). Differences in fear motivated behaviors among inbred mouse strains. *Psychopharmacology, 111*(3), 323-331. doi:10.1007/BF02244948

Tsilidis, K. K., Panagiotou, O. A., Sena, E. S., Aretouli, E., Evangelou, E., Howells, D. W., . . . Ioannidis, J. P. (2013). Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol, 11*(7), e1001609. doi:10.1371/journal.pbio.1001609

van Altena, A. J., Spijker, R., & Olabarriaga, S. D. (2019). Usage of automation tools in systematic reviews. *Research Synthesis Methods, 10*(1), 72-82. doi:10.1002/jrsm.1335

van Luijk, J., Bakker, B., Rovers, M. M., Ritskes-Hoitinga, M., de Vries, R. B., & Leenaars, M. (2014). Systematic reviews of animal studies; missing link in translational research? *PLoS ONE, 9*(3), e89981. doi:10.1371/journal.pone.0089981

Veening-Griffioen, D. H., Ferreira, G. S., van Meer, P. J. K., Boon, W. P. C., Gispen-de Wied, C. C., Moors, E. H. M., & Schellekens, H. (2019). Are some animal models more equal than others? A case study on the translational value of animal models of efficacy for Alzheimer's disease. *Eur J Pharmacol, 859*, 172524. doi:https://doi.org/10.1016/j.ejphar.2019.172524

Vesterinen, H. M., Sena, E. S., Egan, K. J., Hirst, T. C., Churolov, L., Currie, G. L., . . . Macleod, M. R. (2014). Meta-analysis of data from animal studies: a practical guide. *J Neurosci Methods, 221*, 92-102. doi:10.1016/j.jneumeth.2013.09.010

von Elm, E., Poglia, G., Walder, B., & Tramèr, M. R. (2004). Different Patterns of Duplicate PublicationAn Analysis of Articles Used in Systematic Reviews. *JAMA, 291*(8), 974-980. doi:10.1001/jama.291.8.974

Wallace, B. C., Trikalinos, T. A., Lau, J., Brodley, C., & Schmid, C. H. (2010). Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics, 11*. doi:10.1186/1471-2105-11-55

Wallace, B. C., Kuiper, J., Sharma, A., Zhu, M. B., & Marshall, I. J. (2016). Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision. *J Mach Learn Res, 17*.

Walsh, R. N., & Cummins, R. A. (1976). The open-field test: A critical review. *Psychological Bulletin, 83*(3), 482-504. doi:10.1037/0033-2909.83.3.482

Watt, A., Cameron, A., Sturm, L., Lathlean, T., Babidge, W., Blamey, S., . . . Maddern, G. (2008). Rapid reviews versus full systematic reviews: an inventory of

current methods and practice in health technology assessment. *Int J Technol Assess Health Care, 24*(2), 133-139. doi:10.1017/s0266462308080185

Westgate, M. J. (2019). revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods, 10*(4), 606-614.

Wever, K., Ranis, D., Hooijmans, C., & Riksen, N. (2018). The effects of the novel anti-diabetic drugs SGLT2i, GLP-1a and DPP4i on atherosclerosis - A systematic review and meta-analysis of animal studies. *PROSPERO 2018 CRD42018116259*. Retrieved from Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018116259

Wever, K. E., Hooijmans, C. R., Riksen, N. P., Sterenborg, T. B., Sena, E. S., Ritskes-Hoitinga, M., & Warlé, M. C. (2015). Determinants of the Efficacy of Cardiac Ischemic Preconditioning: A Systematic Review and Meta-Analysis of Animal Studies. *PLoS ONE, 10*(11), e0142021. doi:10.1371/journal.pone.0142021

Wexler, Y., Benjamini, Y., & Golani, I. (2018). Vertical exploration and dimensional modularity in mice. *R Soc Open Sci, 5*(3), 180069. doi:10.1098/rsos.180069

Wheater, E. N. W., Stoye, D. Q., Cox, S. R., Wardlaw, J. M., Drake, A. J., Bastin, M. E., & Boardman, J. P. (2020). DNA methylation and brain structure and function across the life course: a systematic review. *Neuroscience & Biobehavioral Reviews*. doi:https://doi.org/10.1016/j.neubiorev.2020.03.007

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. doi:10.1038/sdata.2016.18

Williams, J. H., Li, Y. G., Nayak, A., Errington, M. L., Murphy, K. P. S. J., & Bliss, T. V. P. (1993). The suppression of long-term potentiation in rat hippocampus by inhibitors of nitric oxide synthase is temperature and age dependent. *Neuron, 11*(5), 877-884. doi:https://doi.org/10.1016/0896-6273(93)90117-A

Willner, P. (1984). The validity of animal models of depression. *Psychopharmacology (Berl), 83*(1), 1-16. doi:10.1007/bf00427414

Wiseman, F. K., Al-Janabi, T., Hardy, J., Karmiloff-Smith, A., Nizetic, D., Tybulewicz, V. L., . . . Strydom, A. (2015). A genetic cause of Alzheimer disease: mechanistic insights from Down syndrome. *Nat Rev Neurosci, 16*(9), 564-574. doi:10.1038/nrn3983

Wu, L., Rosa-Neto, P., Hsiung, G. Y., Sadovnick, A. D., Masellis, M., Black, S. E., . . . Gauthier, S. (2012). Early-onset familial Alzheimer's disease (EOFAD). *Can J Neurol Sci, 39*(4), 436-445.

Yamasaki, T., Horie, S., Muranaka, H., Kaseda, Y., Mimori, Y., & Tobimatsu, S. (2012). Relevance of in vivo neurophysiological biomarkers for mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis, 31 Suppl 3*, S137-154. doi:10.3233/jad-2012-112093

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends, 55*(1), 32-36.

Zheng, H., & Koo, E. H. (2006). The amyloid precursor protein: beyond amyloid. *Mol Neurodegener, 1*(1), 5. doi:10.1186/1750-1326-1-5

Zhou, Q., Homma, K. J., & Poo, M. M. (2004). Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron, 44*(5), 749-757. doi:10.1016/j.neuron.2004.11.011

Zimmer, E. R., Parent, M. J., Cuello, A. C., Gauthier, S., & Rosa-Neto, P. (2014). MicroPET imaging and transgenic models: a blueprint for Alzheimer's disease clinical research. *Trends Neurosci, 37*(11), 629-641. doi:https://doi.org/10.1016/j.tins.2014.07.002

Zschenderlein, C., Gebhardt, C., von Bohlen und Halbach, O., Kulisch, C., & Albrecht, D. (2011). Capsaicin-Induced Changes in LTP in the Lateral Amygdala Are Mediated by TRPV1. *PLoS ONE, 6*(1), e16116. doi:10.1371/journal.pone.0016116

Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annu Rev Physiol, 64*, 355-405. doi:10.1146/annurev.physiol.64.092501.114547

Zuroff, L., Daley, D., Black, K. L., & Koronyo-Hamaoui, M. (2017). Clearance of cerebral Aβ in Alzheimer's disease: reassessing the role of microglia and monocytes. *Cellular and Molecular Life Sciences, 74*(12), 2167-2201. doi:10.1007/s00018-017-2463-7