FROM MOUSE TO MEDICINE: OPTIMIZING THE PREDICTIVE VALUE OF PRECLINICAL RESEARCH

One of the major challenges to the successful development of therapies for Alzheimer's disease (AD) is the poor translation of preclinical efficacy from animal models to the clinic. A number of key factors have been identified as contributors to the unsuccessful translation of therapeutic efficacy, these include: the failure of the models to fully recapitulate human AD, poor rigor, study design and data analysis, insufficient attention given to using a standard set of "best practices", failure to match outcome measures used in preclinical animal studies and clinical studies, poor reproducibility of published data and publication bias in favor of reporting positive findings.

To address this challenge and ameliorate some of the factors contributing to the preclinical to clinical gap in the development of AD therapies the National Institute on Aging (NIA) and the National Institutes of Health (NIH) Library have created a publicly available data repository – the **Alzheimer's Preclinical Efficacy Database**, or **AlzPED**. AlzPED is designed as a web-based portal for housing, sharing and mining of preclinical efficacy data. The data are submitted to AlzPED through a curator and gleaned from at least two sources; 1) the scientific literature; 2) directly from researchers. These data include information on AD animal models, therapeutic agents, therapeutic targets, outcome measures, related clinical trials, patents and study design. Most importantly, AlzPED is designed to help identify critical experimental design elements and methodology missing from studies that make them susceptible to misinterpretation and reduce their reproducibility and translational value. Through this capability, AlzPED is intended to guide the development and implementation of strategies and recommendations for standardized best practices for the rigorous preclinical testing of AD candidate therapeutics.

This growing knowledge platform currently houses 917 preclinical efficacy studies published between 1996 and the present (Table 1), collected from databases like PubMed and EMBASE using key word search strings specific to AD. Each study is carefully curated by 2 experts in AD research prior to publication in the database. Efforts are underway to expand the database further and balance the number of studies curated based on the year of publication.

YEAR OF PUBLICATION	# OF PAPERS IN DATABASE	YEAR OF PUBLICATION	# OF PAPERS IN DATABASE
1996	1	2009	39
1999	1	2010	41
2000	14	2011	38
2001	12	2012	59
2002	21	2013	65
2003	13	2014	62
2004	25	2015	76
2005	24	2016	66
2006	50	2017	80
2007	50	2018	75
2008	39	2019	66

Table 1: The table shows the number of preclinical efficacy studies collected, curated and available in the database and the year of publication.

AlzPED identifies 24 experimental design elements that should be included in any preclinical efficacy study to improve its rigor, reproducibility and translational value (Figure 1). Comprehensive analysis of the 917 published preclinical efficacy studies compiled in AlzPED demonstrate considerable variation in the frequency of including and reporting these elements of experimental design. For example, experimental design elements like dose and formulation of the therapeutic agent being examined and treatment paradigms are included and reported with consistency (at 95% or greater) whereas others like power calculation, blinding, randomization, and balancing a study for sex as a biological variable are less frequently reported (less than 35%).
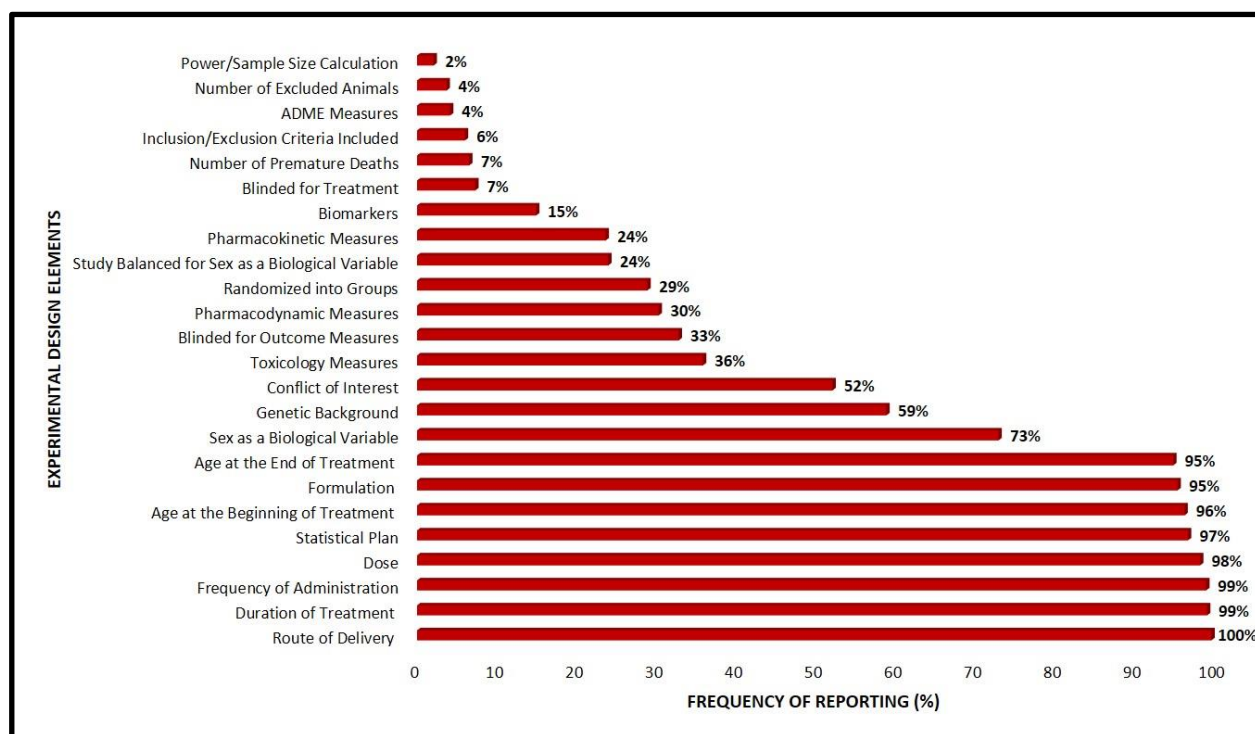

Figure 1: Percentage of studies reporting the 24 experimental design elements identified to improve rigor and reproducibility, calculated from 917 published preclinical studies curated to AlzPED.

AlzPED further defines 9 core experimental design elements that are critical for ensuring scientific rigor and reproducibility of a preclinical efficacy study, derived from Shineman et al., 2011, Landis et al., 2012, Snyder et al., 2016 and ARRIVE guidelines (Figure 2). These include power/sample size calculation, randomization, blinding for treatment allocation and outcomes, sex as a biological variable and balancing for sex as a biological variable, animal genetic background, conflict of interest statement and inclusion and exclusion criteria. Of these 9 core design elements, there has been an upward trend over the past decade in reporting the sex as a biological variable and genetic background of the animals used in the study as well as a financial conflict of interest statement from authors, though this is reflective of changes in data reporting and publication policies and requirements recommended by federal funding agencies, private foundations and peer-reviewed scientific journals. The remaining 6 core design elements including power/sample size calculation, randomization, blinding for treatment and outcomes, balancing for sex as a biological variable, and inclusion and exclusion criteria are poorly reported.
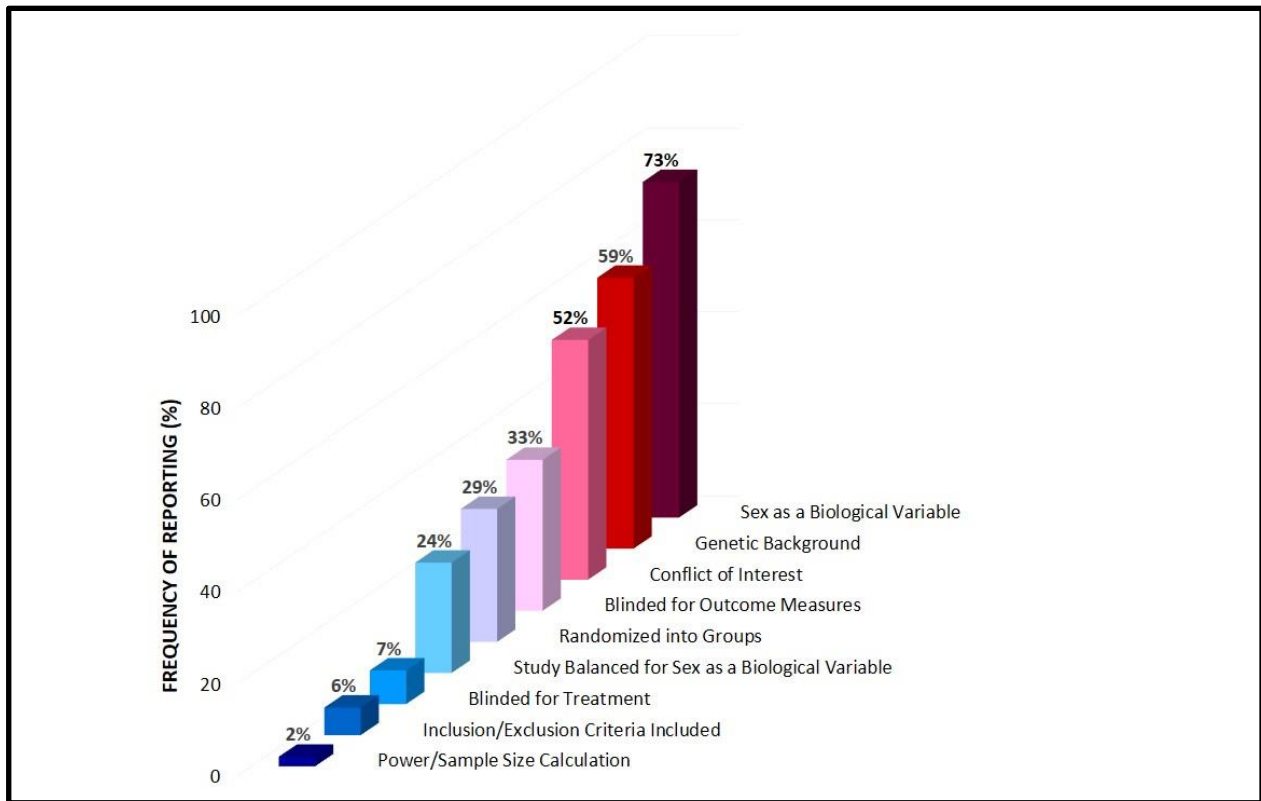
Figure 2: Percentage of studies that reported 9 core experimental design elements derived from Shineman et al., 2011, Landis et al., 2012, Snyder et al., 2016 and ARRIVE guidelines, calculated from 917 published preclinical studies curated to AlzPED.

Further evaluation of the reporting trends in the 9 core experimental design elements demonstrates that few studies report more than 5 core design elements, most studies reporting only 2-4 core design elements. From the 917 curated preclinical studies in the database, 5% report none of the core design elements, 14% report at least 1 core design element, 25% report at least 2, 26% report at least 3, 18% report 4, 8% report 5, 3% report 6, 2% report 7, and 0.2% report all 9 core design elements (Figure 3).
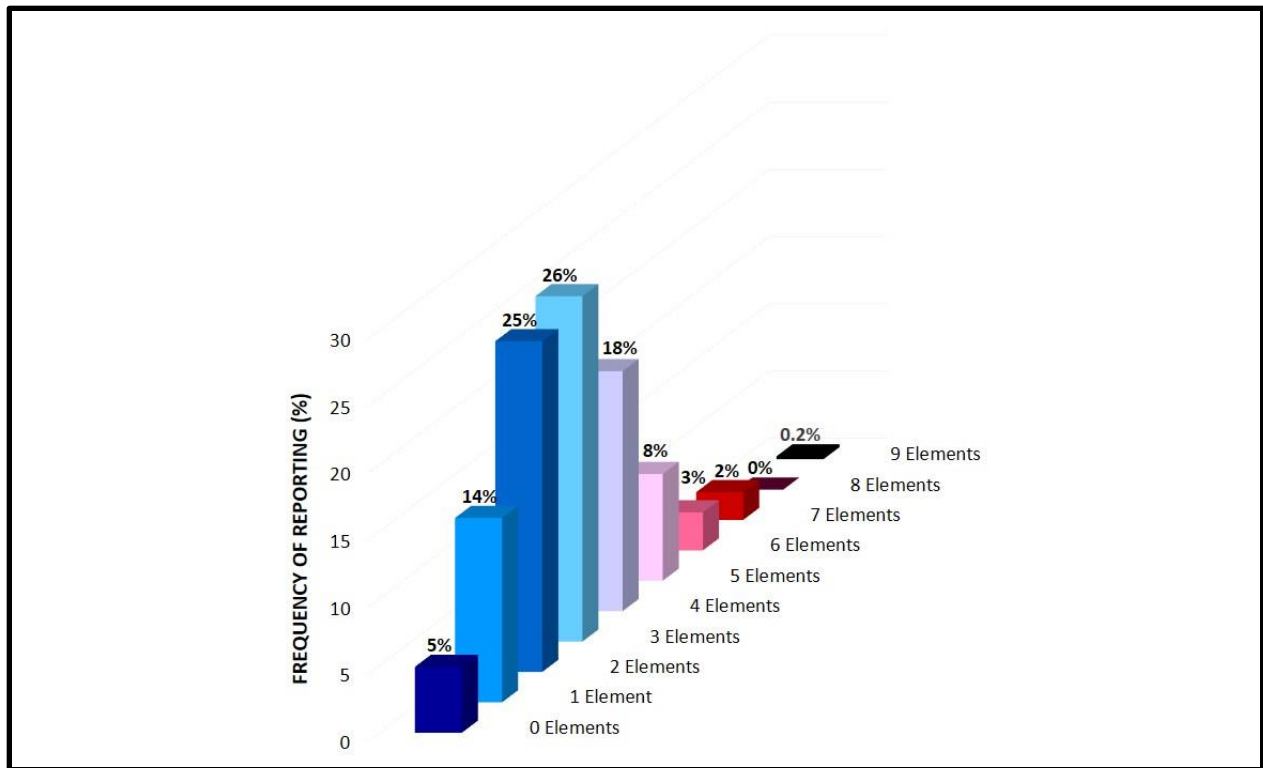
Figure 3: Percentage of studies reporting 0-9 core experimental design elements, calculated from 917 published preclinical studies curated to AlzPED.

Comparisons of the reporting trends in the 9 core experimental design elements between NIH-funded studies and those funded by non-NIH agencies demonstrate considerable variations (Figure 4). Non-NIH funded studies include those funded by nonprofit organizations, pharmaceutical companies, the European Union, and the Chinese, Japanese and Korean governments. NIH-funded studies show a higher frequency in reporting several core design elements, including sex as a biological variable, study balanced for sex as a biological variable, blinding for treatment allocation and outcomes, randomization, and conflict of interest.
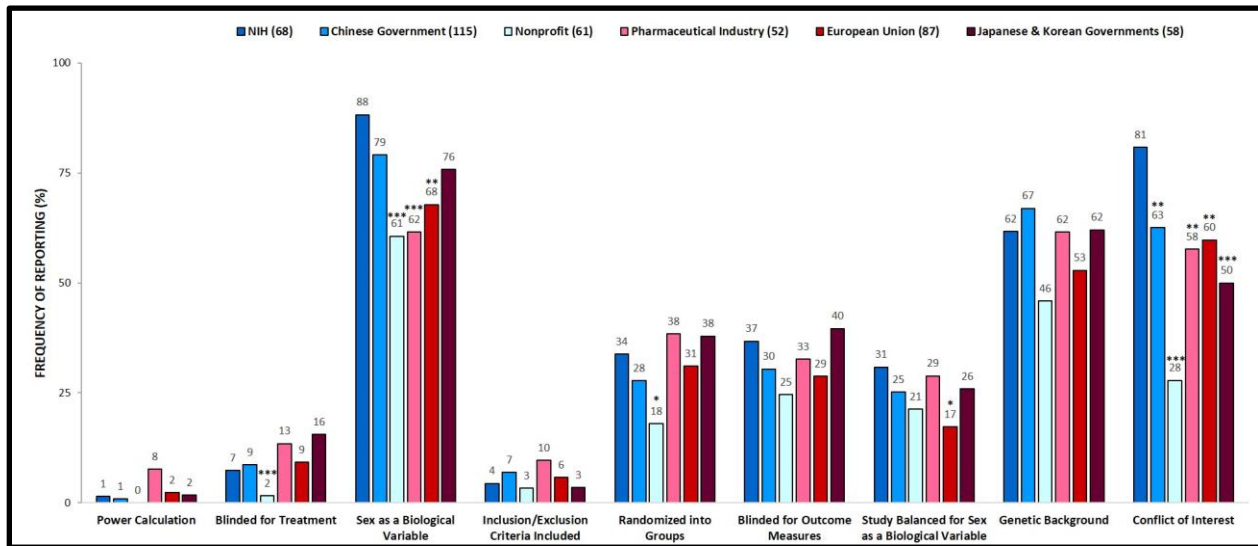
Figure 4: Comparison of reporting trends in 9 core experimental design elements between NIH-funded and non-NIH funded studies. Data are presented as percentages calculated from 68 NIH-funded studies, 115 studies funded by the Chinese Government, 61 Nonprofit-funded studies, 52 Pharmaceutical Industry-funded studies, 87 European Union-funded studies, and 58 studies funded by the Japanese or Korean Governments. Data are analyzed using the Fisher's Exact Test, *p<0.05, **p<0.01 and ***p<0.001.

Reporting trends for the 24 experimental design elements were evaluated based on the impact factor of the journal in which the curated preclinical study was published (Figure 5). These studies were categorized into 4 groups based on 2018 journal impact factor values. Curated studies published in journals with impact factors below 3 were sorted in Group 1, and those published in journals with impact factors between 3 and 4.99, or between 5 and 9.99 were sorted in Groups 2 and 3 respectively. Studies published in high impact journals with impact factors greater than 10 were sorted in Group 4. While t-tests show that there are statistically significant differences in reporting the 24 elements of experimental design between these four groups, overall, the data demonstrate poor reporting practices irrespective of journal impact factor.
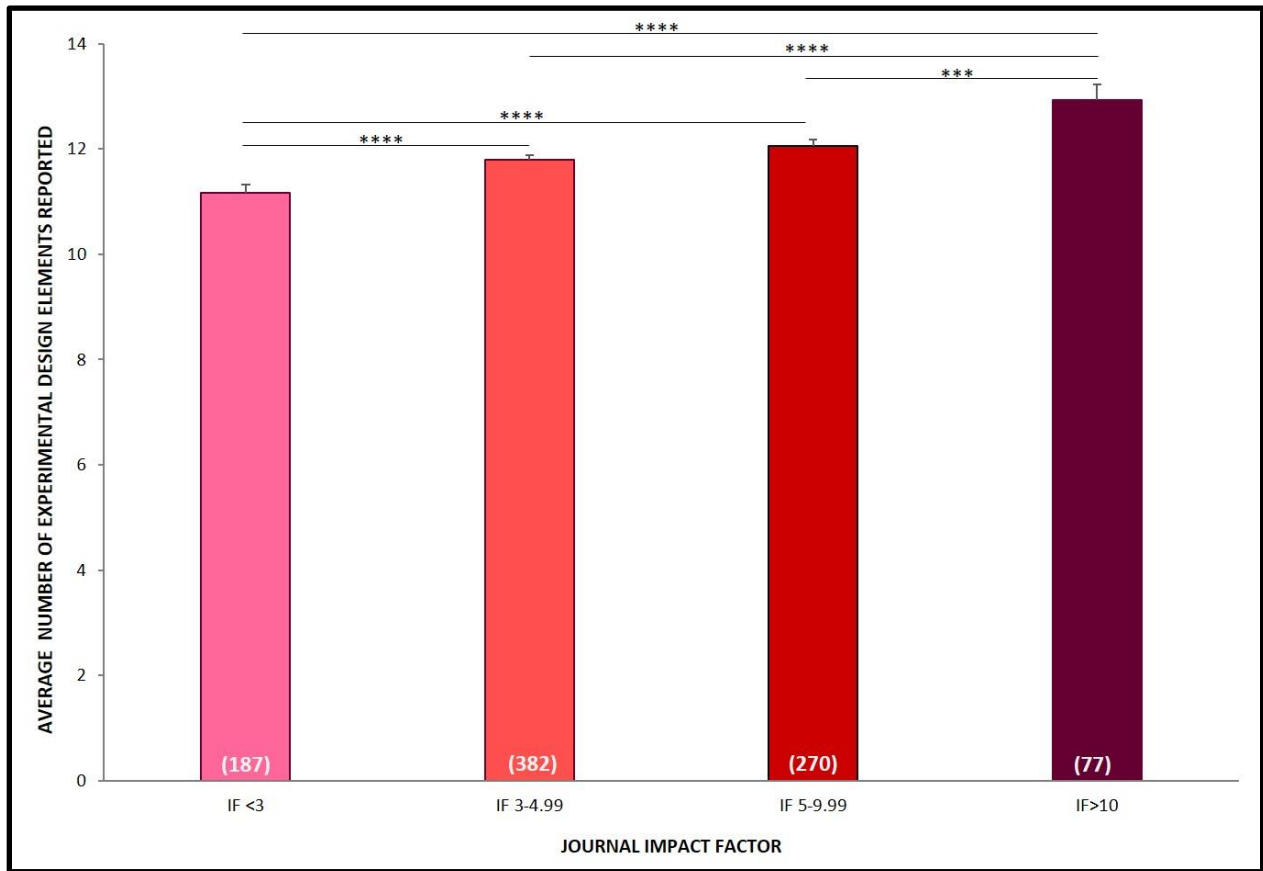
Figure 5: Reporting trends for 24 experimental design elements based on journal impact factor. Data are presented as Mean ± SEM and analyzed using two-tailed t-tests, ***p<0.001 and ****p<0.0001, from 187, 382, 270 and 77 curated studies published in journals with impact factors below 3, between 3 and 4.99, between 5 and 9.99, and above 10 respectively.

Reporting trends for the 9 core experimental design elements were similarly evaluated based on the impact factor of the journal in which the curated preclinical study was published (Figure 6). Curated studies were grouped as detailed in the previous segment. Here as well, even though t-tests show that there are statistically significant differences in reporting the 9 core elements of experimental design between these four groups, overall, the data demonstrate poor reporting practices irrespective of journal impact factor.
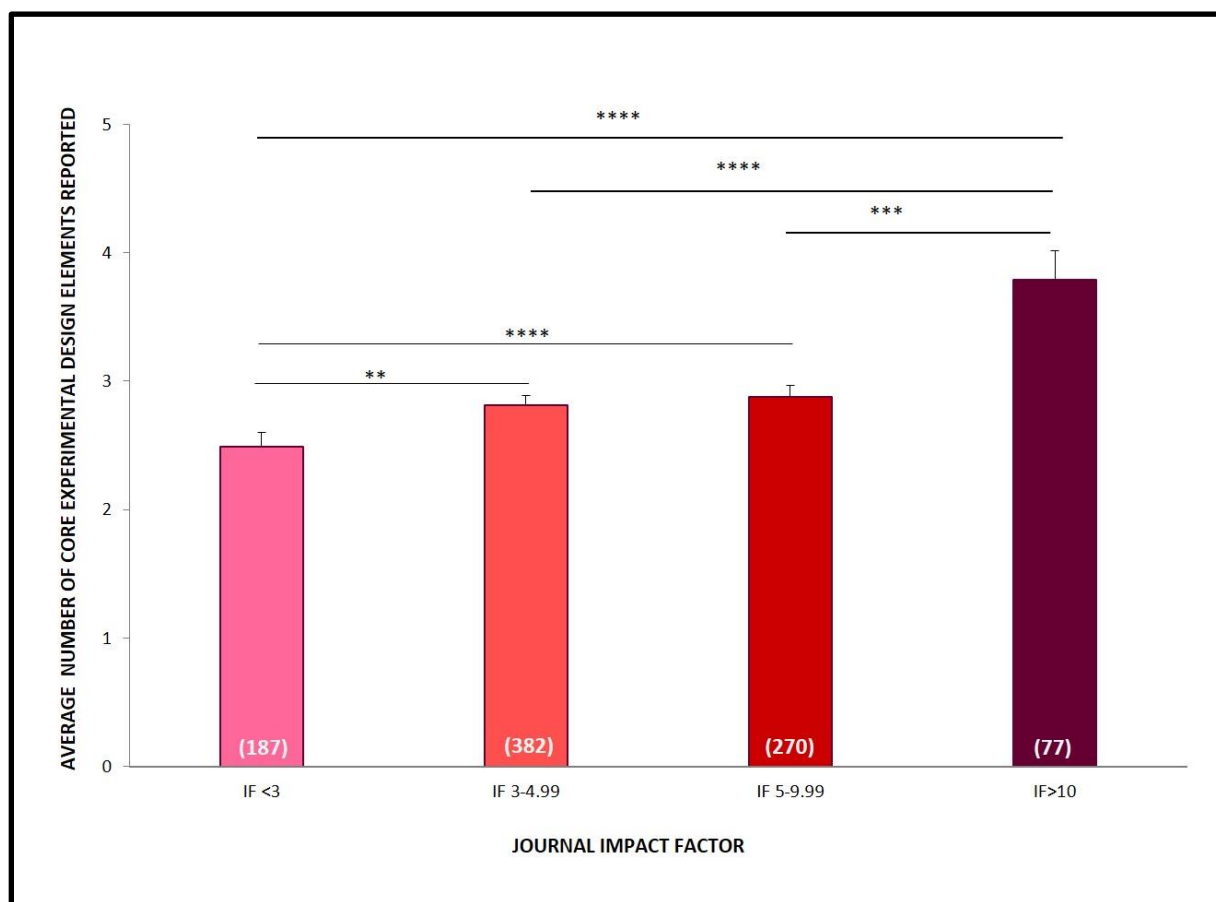
Figure 6: Reporting trends for 9 core experimental design elements based on journal impact factor. Data are presented as Mean ± SEM and analyzed using two-tailed t-tests, **p<0.01, ***p<0.001 and ****p<0.0001, from 187, 382, 270 and 77 curated studies published in journals with impact factors below 3, between 3 and 4.99, between 5 and 9.99, and above 10 respectively.

Reporting trends for the 24 experimental design elements were evaluated based on the relative number of citations per year of each curated study published between 1996 and 2017 (Figure 7). Relative number of citations for each curated study was calculated by dividing the total number of citations for that study by the number of years since publication. For example, for a study published in 2017, the total number of citations for that study was divided by 2, or for a study published in 2016, the total number of citations for that study was divided by 3, and so on. These studies were categorized into 3 groups based on the relative number of citations per year. Curated studies with less than 2.5 relative number of citations per year were sorted into Group 1, those with relative number of citations per year between 2.5 and 7.5 or those with relative number of citations per year greater than 7.5 were sorted into Groups 2 and 3 respectively. While t-tests show that there are statistically significant differences in reporting the 24 elements of experimental design between these three groups, overall, the data demonstrate poor reporting practices irrespective of relative number of citations per year.
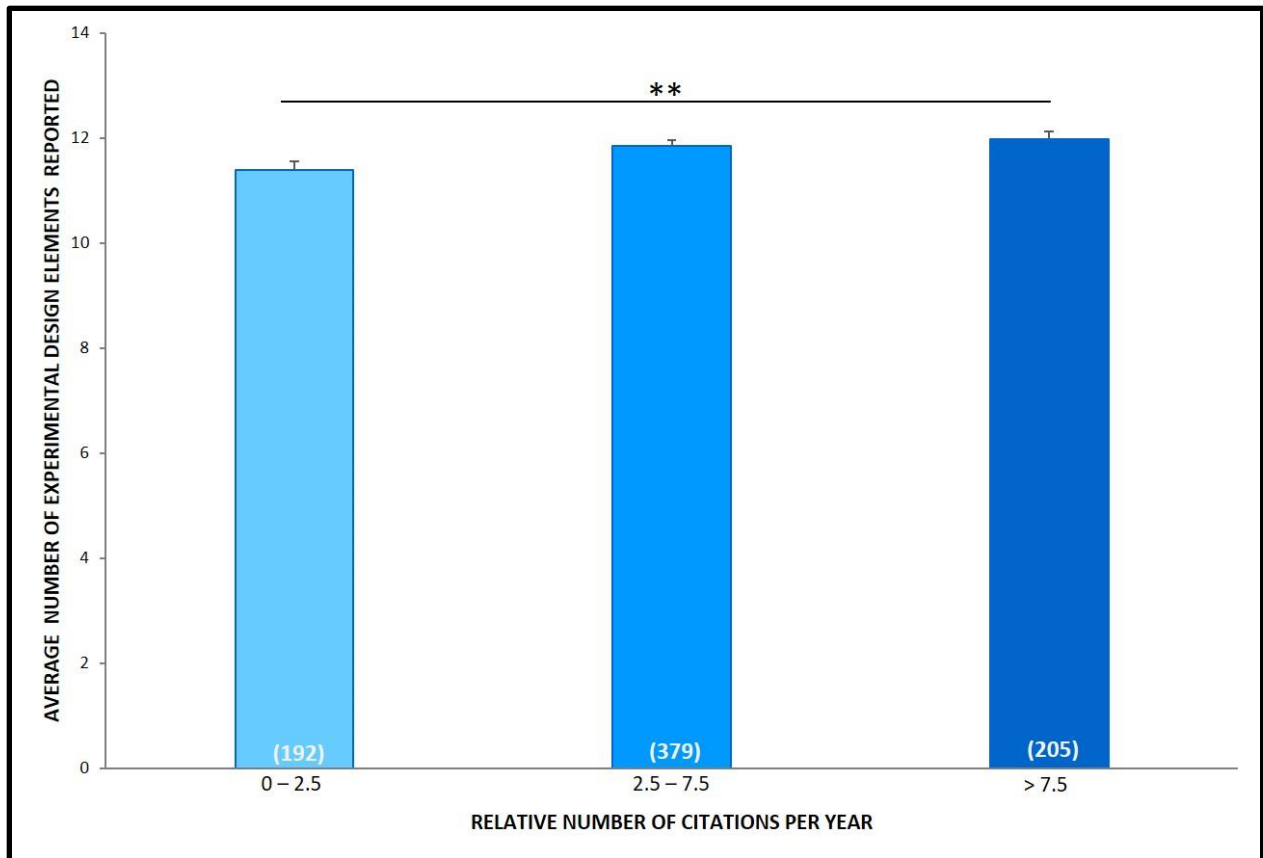
Figure 7: Reporting trends for 24 experimental design elements based on relative citation number per year. Data are presented as Mean ± SEM and analyzed using two-tailed t-tests, **p<0.01, from 192, 379 and 205 curated studies with relative citation numbers per year below 2.5, between 2.5 and 7.5, and above 7.5 respectively.

Reporting trends for the 9 core experimental design elements were similarly evaluated based the relative citation per year of each curated study published between 1996 and 2017 (Figure 8). Curated studies were grouped as detailed in the previous segment. Here as well, the data demonstrate poor reporting of the 9 core elements of experimental design practices irrespective of relative number of citations per year.
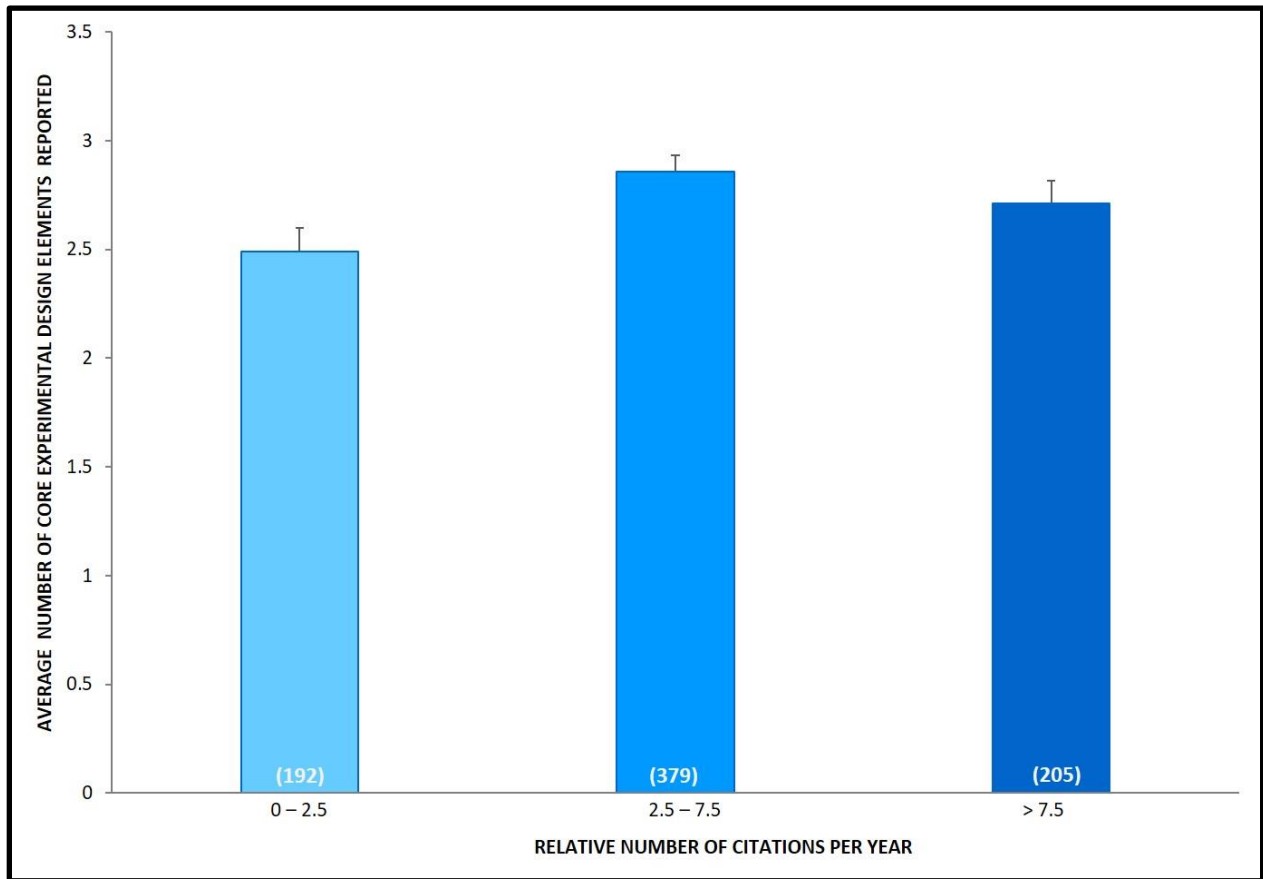
Figure 8: Reporting trends for 9 core experimental design elements based on relative citation number per year. Data are presented as Mean ± SEM and analyzed using two-tailed t-tests, from 192, 379 and 205 curated studies with relative citation numbers per year below 2.5, between 2.5 and 7.5, and above 7.5 respectively.

Within the 917 curated studies compiled in AlzPED, 6 different animal species have been utilized, a majority of which are mouse models of AD (Figure 9). Other animal species include rat, guinea pig, rabbit, dog and non-human primate models of AD. Preclinical efficacy data from 186 different AD animal models are currently available in AlzPED, with new models set to be included as they become available.
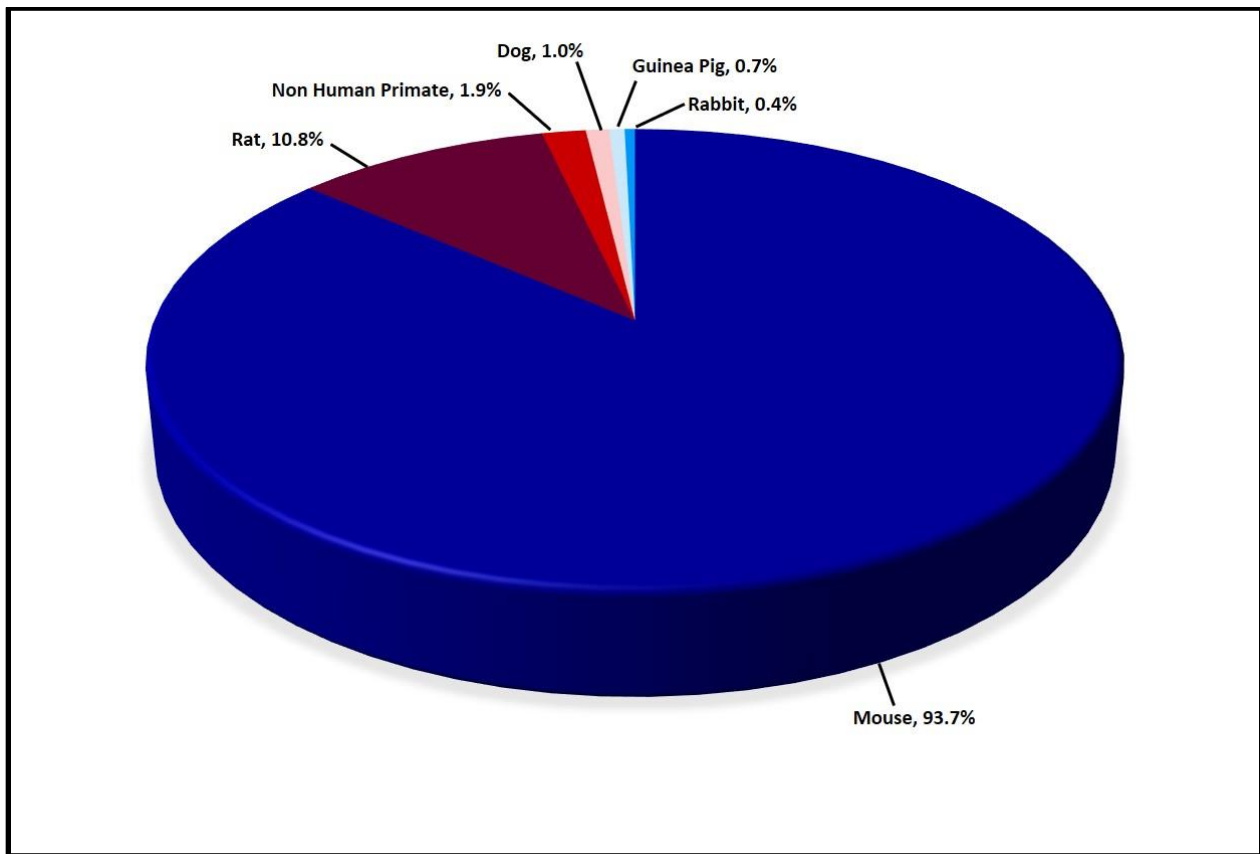
Figure 9: Six different animal species utilized in preclinical efficacy studies, presented as percentages calculated from 917 published preclinical studies curated to AlzPED.

A diverse array of therapeutic agents and targets are reported in the 917 studies curated to AlzPED. The database catalogues 804 novel therapeutic agents into 14 distinct categories (Figure 10) based on agent source (natural product or synthetic), molecular structure (biologic or small molecule), chemical nature (peptide, nucleic acid, or hormone) and mechanism of action (immunotherapy – active or passive).
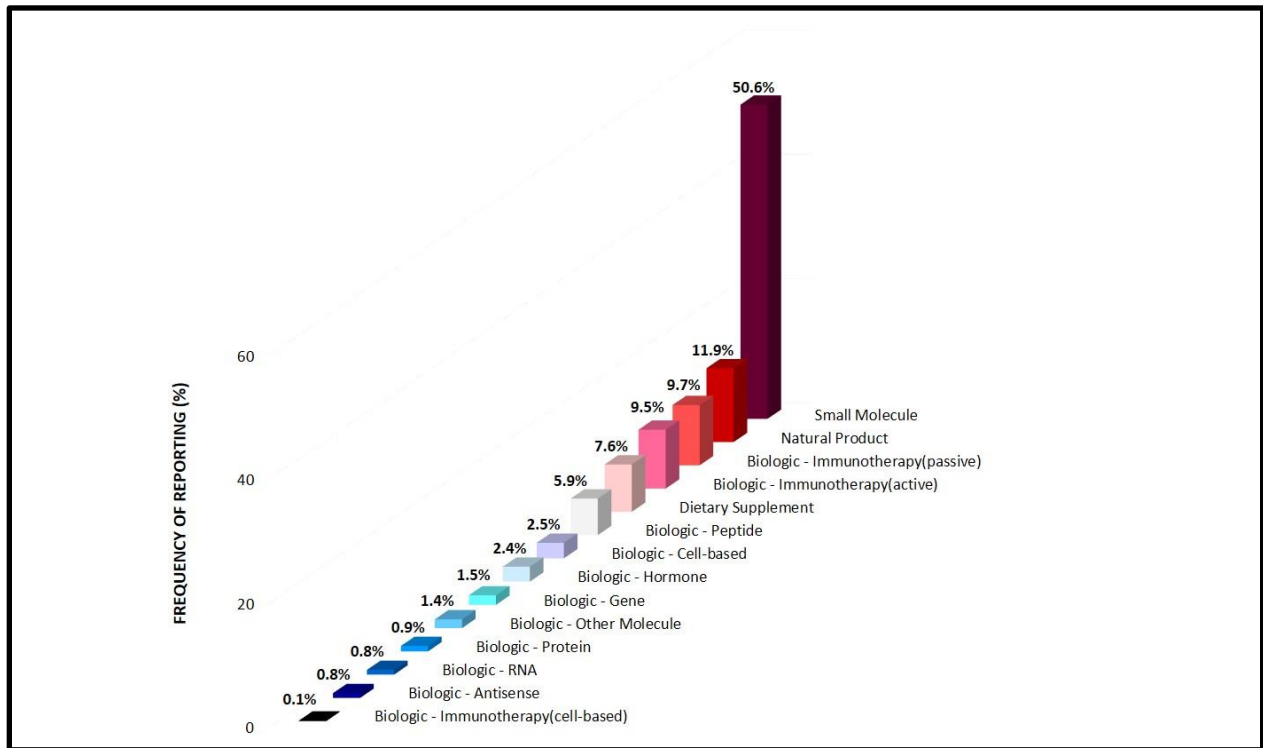
Figure 10: The graph shows the diverse types of therapeutic agents tested, presented as percentages from 917 published preclinical studies curated to AlzPED.

Currently, AlzPED stores information on 175 therapeutic targets that aim to reduce beta amyloid and tau-related pathology and address disease-associated inflammation, oxidative stress, metabolic, synaptic and behavioral dysfunction. These assorted targets are categorized into amyloidogenic proteins, tau protein, non-amyloid proteins, enzymes, receptors and transporters, metal ions, free radicals and multi target (Figure 11).
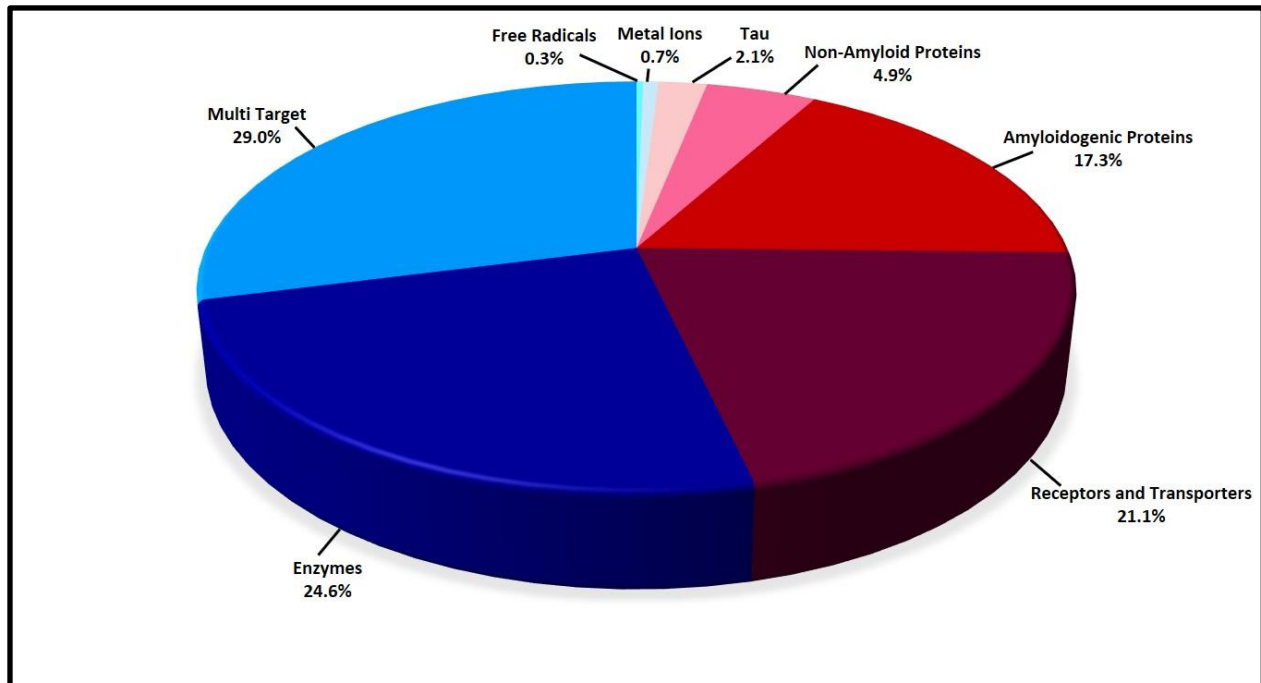
Figure 11: The graph shows the diverse categories of therapeutic targets, presented as percentages from 917 published preclinical studies curated to AlzPED.

Within this diverse group, the most frequently targeted are beta amyloid peptides, beta and gamma secretases, tau protein, cholesterol metabolism regulator HMG CoA reductase, inflammatory response regulating enzyme cyclooxygenase (1 and 2), glucose metabolism regulator peroxisome proliferator-activated receptor gamma (PPAR gamma), and critical neurotransmission and synaptic signaling molecules like NMDA receptors and acetylcholinesterase. Notably, numerous therapeutic agents demonstrate varying extents of anti-inflammatory, anti-oxidant, beta amyloid-reducing, neuroprotective and cognition enhancing properties and are categorized as multi-target therapeutics (Figure 12).
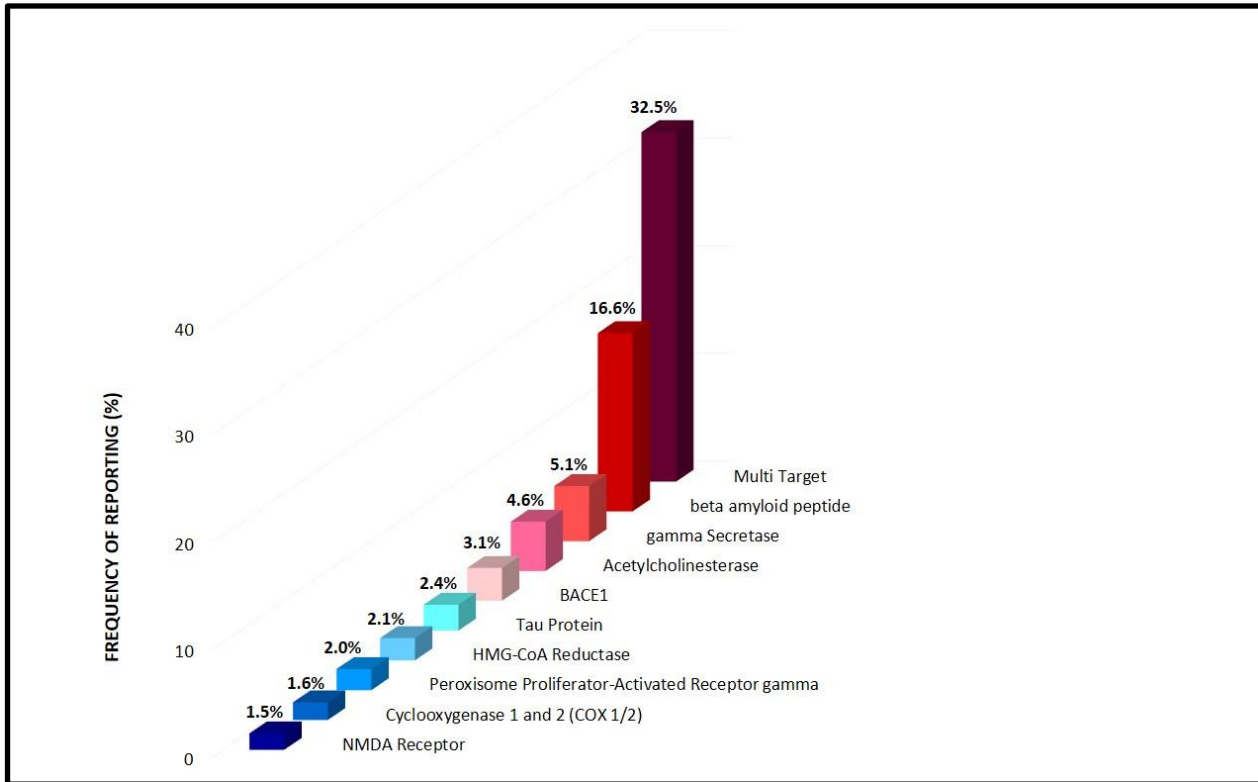
Figure 12: The graph shows the diverse cellular and signaling mechanisms targeted, presented as percentages from 917 published preclinical studies curated to AlzPED.

Each curated study provides an individual snapshot of the measures tested and outcomes achieved in response to the therapeutic agent tested. AlzPED defines 21 different outcome measures that are categorized as either functional or descriptive.

Functional measures include behavioral, motor, electrophysiological and imaging outcomes (Figure 13). Of these functional measures, behavioral outcomes are most commonly tested. There are 73 unique behavioral outcomes measured, from which the Morris water maze, novel object recognition and open field tests are the most frequently studied. Within the 17 different motor function outcomes measured, locomotor activity, swimming speed and the rotarod test are the most frequently studied. 64 diverse electrophysiological outcomes are measured, the most frequently measured being long term potentiation (LTP), field excitatory postsynaptic potentials (fEPSP) and input/output ratio. Within the 38 unique imaging outcomes measured, cerebral blood flow, structural MRI and *in vivo* two-photon amyloid imaging are the most frequently studied
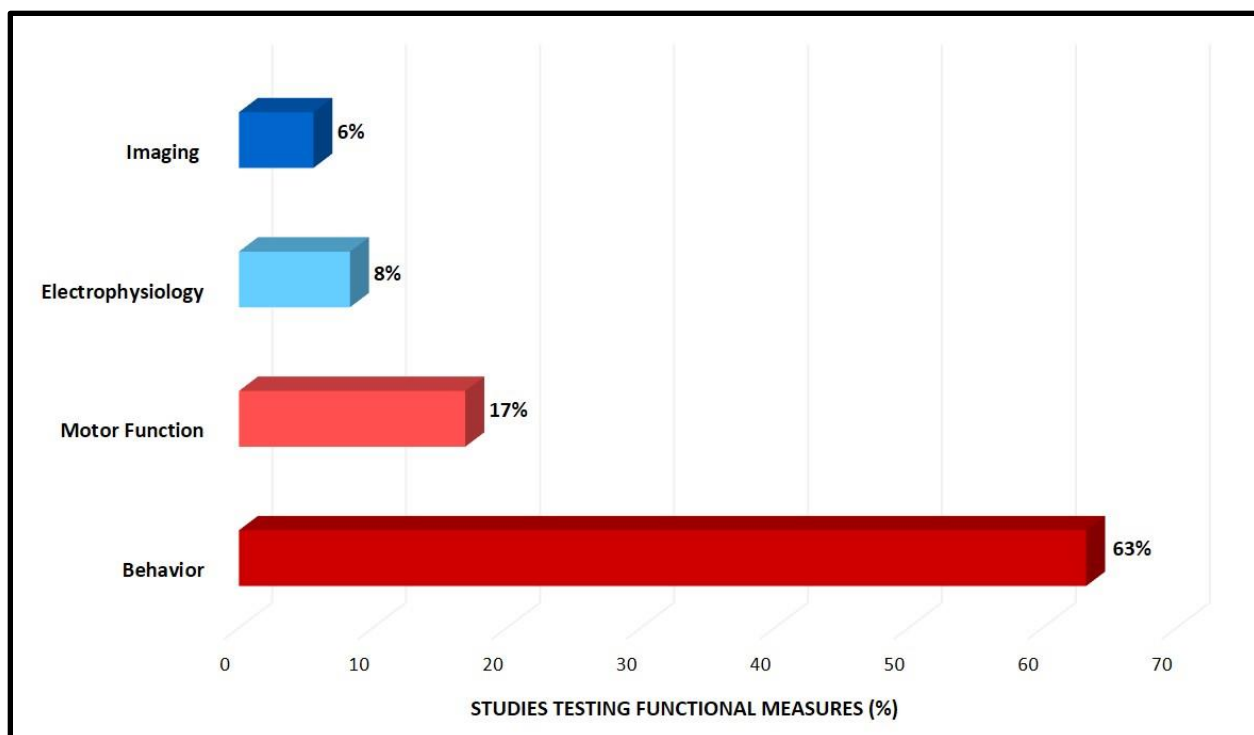
Figure 13: The graph shows the percentage of studies reporting functional measures, calculated from 917 published preclinical studies curated to AlzPED.

Descriptive measures include ADME, biochemical, biomarker, cell biology, chemistry, electron microscopy, histopathological, immunochemical, immunological, microscopy, omics (proteomics, lipidomics, metabolomics, transcriptomics and others), pharmacodynamic, pharmacokinetic, pharmacological, physiological, spectroscopy and toxicology outcomes (Figure 14).

Within the descriptive measures tested, beta amyloid pathology-related biochemical, histopathological, immunochemical and biomarker outcomes are a major focus in the studies curated to AlzPED. These measures analyze several species of beta amyloid including soluble, insoluble, monomers, oligomers, fibrils and plaques. Other measures in these categories include evaluation of several species of tau (soluble, insoluble, aggregated, hyperphosphorylated and others), and astrocytic and microglial markers.

Notably, even though beta amyloid and tau species, and glial markers are a major focus, an extraordinary range of factors and molecules are investigated within these 3 descriptive measures. In total, information from 1184 biochemical, 43 histopathological and 398 immunochemical measures are currently available in AlzPED. As many as 34 different biomarkers have been analyzed, and beta amyloid markers in plasma, serum or CSF constitute a large proportion.
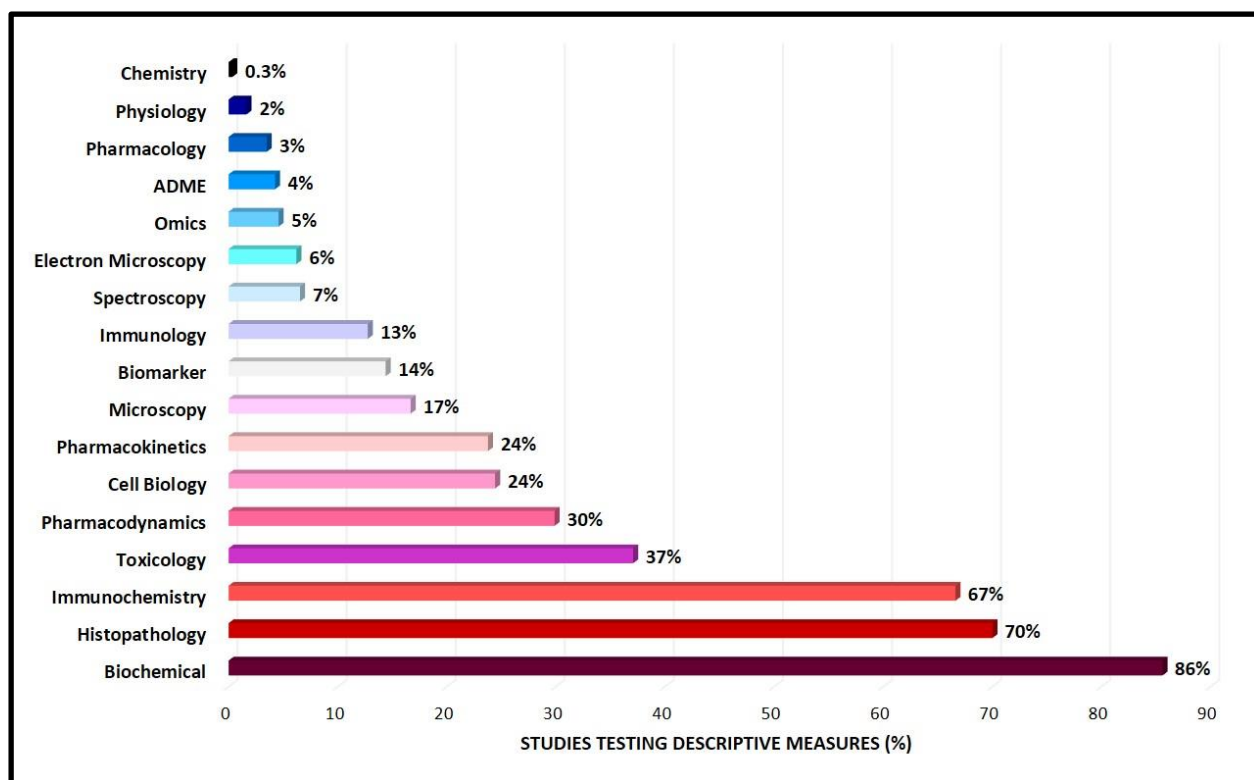
Figure 14: The graph shows the percentage of studies reporting descriptive measures, calculated from 917 published preclinical studies curated to AlzPED.

Other frequently studied descriptive measures used to characterize the therapeutic agent being tested include ADME, pharmacokinetic, pharmacodynamic and toxicology outcomes. Of the 25 ADME measures studied, the most commonly tested are biodistribution, metabolic stability and cytochrome p450 inhibition capability of therapeutic agent. Similarly, 96 different pharmacodynamic measures are examined with key focus on reducing beta amyloid species. As many as 55 pharmacokinetic measures have been analyzed, and drug concentration in brain and plasma are most frequently evaluated. A comprehensive listing of at least 93 toxicology measures such as Ames tests, enzyme profiles, organ histology and others are available in the database as well. Of these, the most frequently evaluated are body weight, general behavior and food intake.

AlzPED reports on 14 different physiological measures from which blood pressure and cerebral blood flow are most frequently evaluated and 7 pharmacological measures from the most commonly tested are binding affinity and target selectivity of the therapeutic agent. As many as 90 cell biology outcomes are measured, and cell viability and cytotoxicity are the most common measures. Of the 34 immunological measures reported, antibody titers and target specificity are most frequently evaluated.

AlzPED also informs on 13 OMICS-related measures such as metabolomics and gene expression profiles. Finally, AlzPED also reports on 34 electron microscopy outcomes, 55 microscopy outcomes and 14 spectroscopy outcomes.

Analysis of curated studies in AlzPED demonstrates serious deficiencies in reporting critical elements of methodology such as power/sample size calculation, blinding for treatment/outcomes, randomization, sex of animal used and balancing for sex, animal genetic background and others. These deficiencies in reporting critical elements of methodology is also is demonstrated in high impact factor journals as well as highly cited published preclinical research. These deficiencies in study design and methodology diminish the scientific rigor, reproducibility and translational value of the preclinical studies, making it evident that a standardized set of best practices is required for successful translation of therapeutic efficacy in AD research.

In summary, AlzPED provides the AD research community free access to a treasure trove of information pertaining to experimental designs, Alzheimer's animal models, therapeutic agents and targets, outcome measures, and principal findings from preclinical studies, along with related PubMed and PubChem literature, clinical trials, patents, funding sources and financial conflict of interest. AlzPED is also designed to serve as a platform for reporting unpublished negative findings to mitigate publication bias favoring reporting of positive findings. Researchers can use this resource to survey existing preclinical therapy developments, understand the requirements for rigorous study design and transparent reporting and plan preclinical intervention studies.

<u>**REFERENCES CITED:**</u>

1. Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. Shineman et al., Alzheimers Res Ther. 2011, Sep 8;3(5):28.
https://www.ncbi.nlm.nih.gov/pubmed/21943025

2. A call for transparent reporting to optimize the predictive value of preclinical research. Landis et al., Nature. 2012, Oct 11;490(7419):187-91.
https://www.ncbi.nlm.nih.gov/pubmed/23060188

3. Guidelines to improve animal study design and reproducibility for Alzheimer's disease and related dementias: for funders and researchers. Snyder et al., Alzheimers Dement. 2016, Nov;12(11):1177-1185.
https://www.ncbi.nlm.nih.gov/pubmed/?term=27836053

4. ARRIVE Guidelines - https://www.nc3rs.org.uk/arrive-guidelines