AlzPED – Rigor, Reproducibility, Transparency

Alzheimer's Preclinical Efficacy Database

Cindy Sheffield, MBA, MLS
Alzheimer's Preclinical Efficacy Database
Zimmerman Associates, Inc.
Fairfax, Virginia

Abstract

The Alzheimer Preclinical Efficacy Database (AlzPED) has been developed by the National Institute on Aging and the National Institutes of Health Library. It focuses on improving reporting standards of animal studies, and collecting both published and unpublished research in a collective repository. This presentation will demonstrate how expert knowledge in Alzheimer's disease (AD) therapeutic drug discovery and technical information were combined for the effective design of AlzPED. Scientists' envisioned a database to facilitate discovery and improve awareness about rigor and reproducibility in AD research. They hired a librarian to help.

Experiment attributes that demonstrate best practices of the scientific method were identified. These attributes were organized into logical sections to facilitate comprehension, information flow, and comparisons across studies. A score card approach has been used to identify which scientific elements have been reported within each study, and the overall assessment provides an indication of rigor and reproducibility. Links to external resources allow for deeper learning and discoverability. This system compliments information found within PubMed.

The AlzPED's system is intended to improve investigators' understanding of relationships between animal models, therapeutic targets, therapeutic agents, and measurable outcomes for AD research. AlzPED has been beta tested and released for use. Results from beta testing support the need ontologies to clearly identify entities within the research. Early comments from professionals in the field have been positive. This talk will review feedback from beta testers, and show early usage statistics.

The target audience for this body of knowledge is very specific, and focuses on those who study neurodegeneration of the brain due to dementia. The critical nature of a cure for this disease supports the effort to standardize terminology, improve rigor, and increase transparency. These efforts, along with scientific advances, may increase the likelihood for effective treatments for Alzheimer's disease.

Introduction

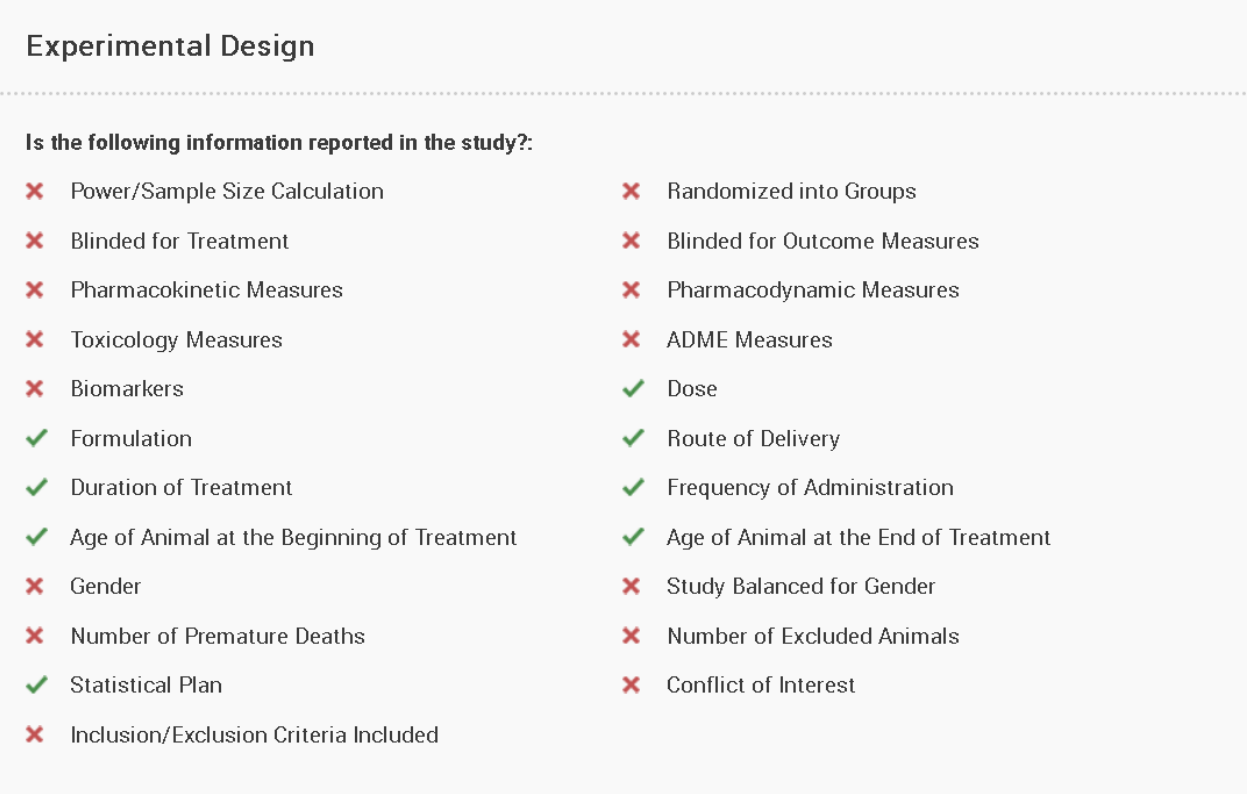
Program Directors at the National Institute on Aging want to increase reproducibility in Alzheimer's drug development studies, and ensure funding goes toward studies with maximum effectiveness. They worked with the National Institutes of Health Library to develop the International Alzheimer's Disease Research Portfolio, which allows funding organizations to quickly assess who is funding what projects. Based on this success, NIA and the NIH Library teamed up to develop the Alzheimer's Disease Preclinical Efficacy Database. This paper will discuss the need to promote rigor and reproducibility, the identification of ontologies to unify subject areas, and the early feedback received for the website.

Rigor and Reproducibility

PubMed does an excellent job of indexing literature, but specific research areas with hard to treat diseases, such as Alzheimer's disease, can benefit from specific reporting requirements. NIA Program Directors identified several key fields they wanted to include within AlzPED to focus on reporting of research to improve rigor and reproducibility specific to Alzheimer's treatment studies(1). The ARRIVE guidelines(2) and other sources were reviewed before a final selection of fields was made for AlzPED. The data selected fell into five general areas: Bibliographic, Therapeutic Agent, Animal Model, Experimental Design, and Outcomes.

The Experimental Design section was a key area for determining the rigor and reproducibility of many studies(3, 4). Originally the goal was to collect specific values for many of the experimental design elements, such as the volume and frequency agents were administered and methods used to administer those agent. However after reviewing a sample set of 35 documents it became apparent study designs can vary widely and it was going to be difficult to develop a database structure to capture all of the potential variations in study design.

It was determined that knowing whether data was reported or not reported within a study was helpful, and provided enough information for someone to determine whether they wanted to view the full article or report. Thus we came up with the idea of offering checkboxes that would display as checks and X's. (See figure below.)



Almost every field on that screen should have a green check mark. Of the 267 citations curated to date, on average only half of the fields have green checks. The majority of citations within AlzPED show a red X for the 'Power/Sample Size Calculation'. This is of significant concern, as each study should have Power Calculation. If you do not know the power calculation you really cannot determine the validity of the study.

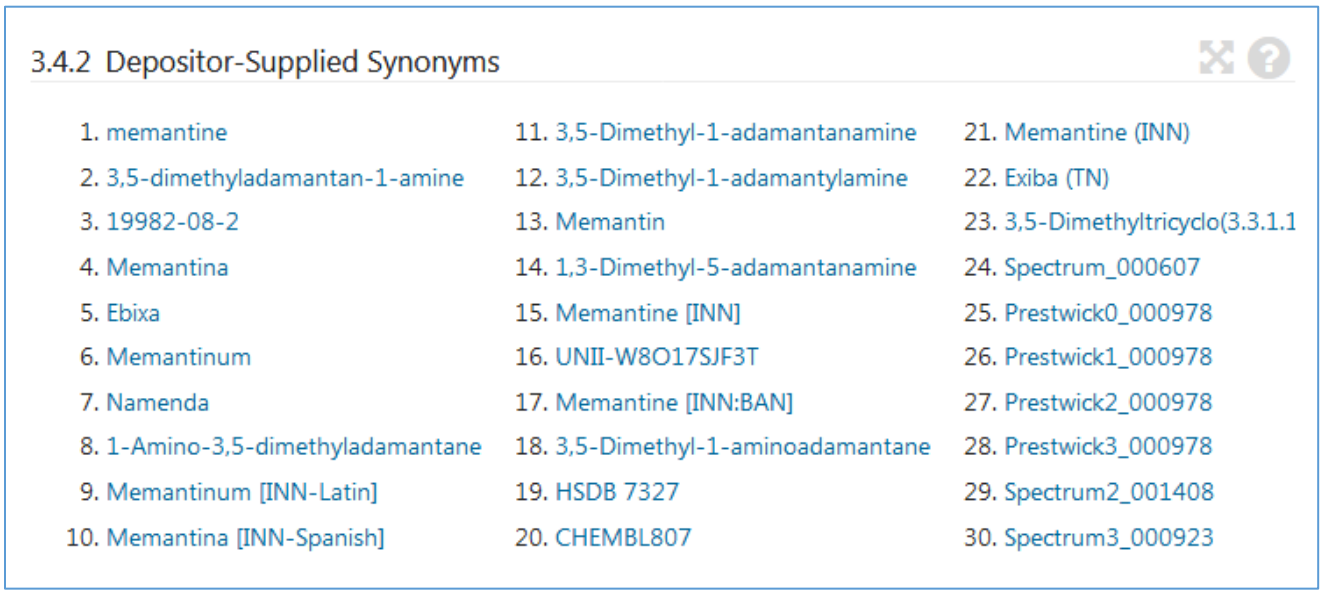
Experiment Attribute	Ratio	Percentage
Power/Sample Size Calculation	1/267	.37%
Blinded for Treatment	26/267	9.73%
Formula	248/267	96.87%
Duration of Treatment	258/267	92.88%
Number of Premature Deaths	21/267	8.17%
Inclusion/Exclusion Criteria	1/267	.37%

Ontologies

The disparate terms used for naming entities within AD scientific experiments demonstrates the need for more unified ontologies around key subject areas. These areas included: Therapeutic Agents, Therapeutic Targets, Animal Models, and Outcomes Measures. This lack of clearly defined terms, makes it difficult for scientists to identify, compare and evaluate information across studies.

Therapeutic Agents

The International Union of Pure and Applied Chemistry offer many rules for naming chemical compounds. There is the Blue Book, Red Book, Green Book, Gold Book, White Book and Orange Book each detailing methodologies and explanations for naming a variety of compounds. Despite these guidelines, identifying various compounds within the literature can be tricky. Compounds can be referenced using their chemical names, which may be punctuated in a variety of ways. Drugs future along on the development pipeline can be called by their generic and brand names. To illustrate the Depositor-Supplied Synonymy's from PubChem for an established medication, Memantine, is shown below.



3.4.2 Depositor-Supplied Synonyms

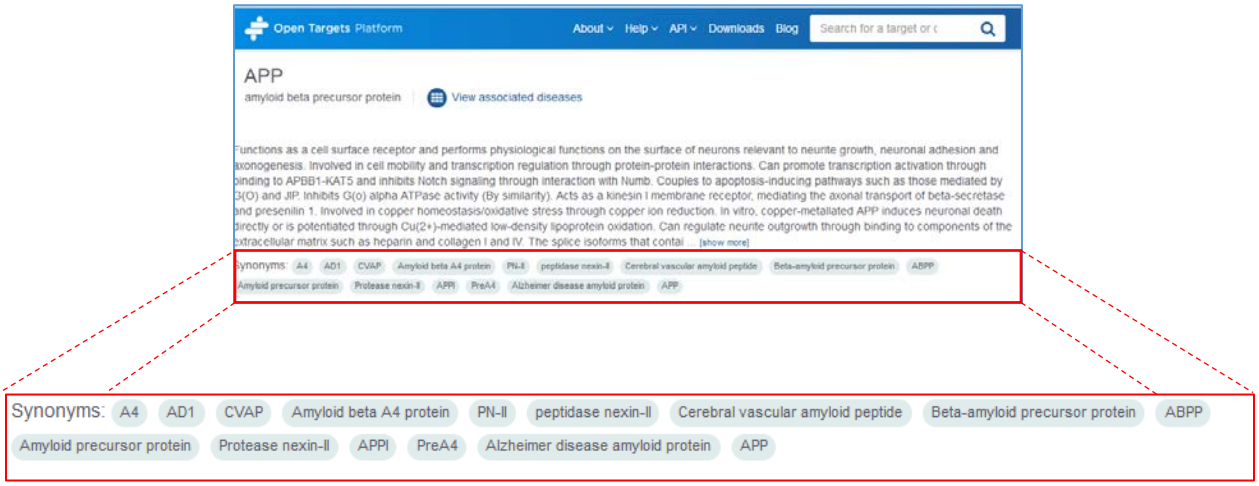
1. memantine
2. 3,5-dimethyladamantan-1-amine
3. 19982-08-2
4. Memantina
5. Ebixa
6. Memantinum
7. Namenda
8. 1-Amino-3,5-dimethyladamantane
9. Memantinum [INN-Latin]
10. Memantina [INN-Spanish]
11. 3,5-Dimethyl-1-adamantanamine
12. 3,5-Dimethyl-1-adamantylamine
13. Memantin
14. 1,3-Dimethyl-5-adamantanamine
15. Memantine [INN]
16. UNII-W8O17SJF3T
17. Memantine [INN:BAN]
18. 3,5-Dimethyl-1-aminoadamantane
19. HSDB 7327
20. CHEMBL807
21. Memantine (INN)
22. Exiba (TN)
23. 3,5-Dimethyltricyclo(3.3.1.1
24. Spectrum_000607
25. Prestwick0_000978
26. Prestwick1_000978
27. Prestwick2_000978
28. Prestwick3_000978
29. Spectrum2_001408
30. Spectrum3_000923

Therapeutic Targets

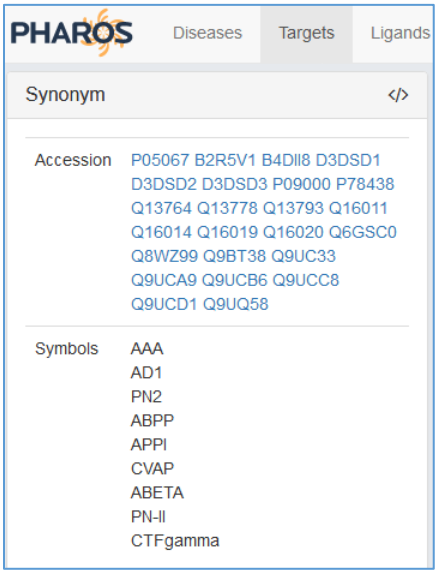
Therapeutic Targets offer their own set of ambiguous terms. For instance one of the primary targets for the treatment of Alzheimer's disease is "amyloid beta precursor protein". It is often referred to as APP, but it is referenced as a variety of other terms, abbreviations, and acronyms

as well. It is called Abeta protein, αβ protein, simply αβ, Amyloid beta, Amyloid β, αβ peptide, and on and on. Just to demonstrate the depth and breadth of the problem here are the listed synonyms for amyloid beta precursor in two open source resources that provide information about therapeutic targets: Open Targets(5) and Pharos(6). Pharos is the Knowledge Management Center for the Illuminating the Druggable Genome, funded by the National Institutes of Health.

Open Targets



Pharos



Animal Models

Animal models can be procured from laboratory animal model providers. Charles River, Jackson Labs, and Taconic Biosciences are a few of the well-known suppliers of AD research mouse models. Other emerging models are produced in laboratories and validated for their genetic background and pathophysiological properties. Two examples of this include Eliezer Masliah's Thy1-hAPP751 mice model and Thomas Bayer's APP/PS1KI mice model. Animal models are often described by their cell line and genetic background, but as with therapeutic agents and targets, there are multiple ways of describing cell lines and genetic grounds. These variations create obstacles in identification and discovery.

AlzForum, (7) operated by the Biomedical Research Forum, provides a news website and information resource dedicated to helping researchers accelerate discovery. Within the site they offer an excellent database for describing each animal mode and provide a list of known synonyms for the model. They also provide genetic background and strain information.

Below is sample data from AlzForum:

Model Name	Synonyms	Description
mThy1-hAPP751 (TASD41)	**Line 41** **hAPPSL** **hAPP-SL** **AβPP751** **mThy1-hAβPP751 Swe Lon (line 41)** **APP751SL** **hAPPlon/swe line 41** **APP41**	Strain Name: **mThy1-hAβPP751 Swe Lon** Genetic Background: **C57BL/6 x DBA**
APP751SL/PS1KI	**APP(SL)PS1KI** **APPxPS1-Ki** **APPSL/PS1KI** **APP(SL)/PS1(KI)** **APP/PS1KI**	Strain Name: **N/A** Genetic Background: **The PS1KI line was established in 129SV and backcrossed >7 times to C57BL/6 background. The PS1KI were bred with APPSL mice on a C57BL background (two rounds) to obtain a homozygote PS1KI and heterozygote APP.**

Outcomes

Outcomes data is some of the most difficult information to assess and synthesis for analysis. Prior to AlzPED there was not a clear definitive method for describing outcomes data in AD therapeutics studies. There are a variety of terms used, some meaning the same things, other meaning very similar, yet different.

Dr. Refolo categorized Outcome Measures using a two-tier approach. First he divided outcomes into 19 different broad categories. These categories include:

Behavioral	Electron Microscopy	Electrophysiology
Motor Function	Spectroscopy	Pharmacokinetics
Histopathology	Imaging	Pharmacodynamics
Biochemical	Cell Biology	Toxicology
Immunochemistry	Immunology	ADME (absorption, distribution, metabolism, and excretion)
Microscopy	Biomarker	Pharmacology
		OMICS

Within each of these broad categories, Dr. Refolo further refined outcomes by using specific Outcome Parameters found within the AD therapeutic discovery literature. The number of Outcome Parameters within each Outcome Measure can vary. For instance, within the category of Motor Function there are nine measurable Outcome Parameters, whereas the Biochemical section offers 294 Outcome Parameters.

Specific Outcome Parameters have been narrowed down from a variety of terms as well. For instance: phosphor-Tau is also referred to as:

Tau Phosphorlation or Phosphorylated Tau

There are an abundance of terms within the Alzheimer's disease literature with similar scenarios. Using this as an example, this is a category of Immunochemistry may have started with 140 terms, but has been pared down to 101 terms that most accurately describe the science within the literature.

Future Directions

Additional work will be can be done with these synonyms in the future by using them to text mine the literature to identify studies for AlzPED. Additionally these properly mapped terms can be used to enhance discovery within AlzPED by allowing researchers to select either a range of terms, or more specific terms, or associated terms. This improvement within discovery alone may help improve rigor and reproducibility.

Early Feedback

AlzPED is still relatively early in its development and acclimation within the Alzheimer's disease research community. Concerted efforts to develop exposure to the database are just beginning. A few presentations have been made to communities such as this, and to the Neuroscience Community. Feedback about AlzPED has been collected by conducting beta tests, looking at usage statistics.

Beta Testing

The overall initial design of AlzPED was completed in December of 2015. Data was entered into AlzPED and beta testing was conducted during the spring of 2016. We were interested in learning views on: the organization of information, the navigation of the tool, the value of the content, and opinions on whether this database would help in making informed research decisions. Three groups of beta testers were interviewed: NIH Librarians, University of Maryland Health Science Librarians, and Leaders from Alzheimer's funding organizations. Both positive and negative feedback was received. Much of the negative feedback was related to the difficulty in discovery related to the synonyms discussed in the previous section. Comments included:

Things that need to be improved

- "A glossary of terms would be really helpful" That might placate some of these [discoverability/variability in results] issues. Perhaps employ a "Consider Using these Terms".
- "Findability – Sample searches: "3xTg" = 18 hits; "3xTg-AD" = 10 hits; "3xTgAD" = 1 hit; "triple transgenic" = 6 hits; "APPxPS1xTau" = 17 hits"
- "Variability in results is problematic"
- "No internal controlled vocabulary to pick up synonyms; users will become frustrated as a result because it will not be usable."
- "Information is a little tricky to be discoverable by a general researcher. For instance, if they're utilizing a specific term or abbreviation (ex. ABP), only one result will come up. However, if they typed in Amyloid beta Peptides, they'll have 92 results."
- "Searchers may want to filter results according to the Quality Measures in the Experimental Design section."
- "Curious about the ease of making edits and what that process is. Not knowing that might make me reluctant to add data to the repository."

Things that have been done well

- "Very helpful in allowing investigators to take a quick look to see what is out there, see what work could be done in-house, and assess what work could be done more quickly. It allows the investigator the ability to assess more accurately what resources need to be brought to the investigation in terms of time and budget."
- "Love what you are doing; providing the ability to drill down to the disease; assist translational research; highlight key elements. Disconcerting how my own publication rated in the assessment. I had some of that information and did not include it in the publication."
- "It will change the culture when people have to enter their own studies, and they know they have to address all of these issues [Experiment Design]."
- "This offers one less step of searching which is nice for someone reviewing studies in the discipline or collecting information."

- "A great site for preclinical models as long as scientists populate it. It is easy to navigate, has a lot of functionality and is easy to upload data. The search function was fantastic."

An announcement about AlzPED was posted from the National Institute on Aging, "Inside NIA: A Blog for Researchers" in January of 2017. That announcement generated traffic of 111 users who came to explore the site. 12 users took the time to register for an account. 3 users visited from Great Britain, and one was from Canada. Additionally, AlzPED was presented during the NIH Policy and Evaluation Fellows Meeting. This particular group is interested in learning about the rigor in preclinical experiments overtime so we will likely be working with this group to explore that.

Currently we are looking at methods for getting AlzPED into the investigators' workflow. We are working with the National Library of Medicine's LinkOut Team to have PubMed articles link to AlzPED articles. We are also looking at incorporating a variety of comments from other services, so that comments about a specific research study can be aggregated within AlzPED.

Conclusion

AlzPED is new and different database design specifically for Alzheimer's disease investigators. It minimizes the duplication of information provided by PubMed, while allowing investigators the ability to quickly drill down to animal model, therapeutic target, therapeutic agent, and outcomes information. The intention of this database is to allow researchers to quickly assess what experiments have been done, and what resources they may want to include in their next set of experiments.

Acknowledgements

I'd like to acknowledge the AlzPED Team Lorenzo M. Refolo, Suzana S. Petanceska for their conception and initiation of this project, and Katerina Mancevska and Zane Martin for all of their hard work and import into the database.

Bibliography

1. Refolo L, Petanceska, S., King, J. Statement of Work for the Alzheimer's Preclinical Efficacy Database. 2014.

2. Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG. Animal research: reporting in vivo experiments: the ARRIVE guidelines. British journal of pharmacology. 2010;160(7):1577-9. Epub 2010/07/24. doi: 10.1111/j.1476-5381.2010.00872.x. PubMed PMID: 20649561; PubMed Central PMCID: PMCPMC2936830.

3. Shineman DW, Basi GS, Bizon JL, Colton CA, Greenberg BD, Hollister BA, et al. Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. Alzheimer's research & therapy. 2011;3(5):28. Epub 2011/09/29. doi: 10.1186/alzrt90. PubMed PMID: 21943025; PubMed Central PMCID: PMCPMC3218805.

4. Snyder HM, Shineman DW, Friedman LG, Hendrix JA, Khachaturian A, Le Guillou I, et al. Guidelines to improve animal study design and reproducibility for Alzheimer's disease and related dementias: For funders and researchers. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2016;12(11):1177-85. Epub 2016/11/12. doi: 10.1016/j.jalz.2016.07.001. PubMed PMID: 27836053.

5. Reisdorf WC, Chhugani N, Sanseau P, Agarwal P. Harnessing public domain data to discover and validate therapeutic targets. Expert opinion on drug discovery. 2017. Epub 2017/05/13. doi: 10.1080/17460441.2017.1329296. PubMed PMID: 28494630.

6. Nguyen DT, Mathias S, Bologa C, Brunak S, Fernandez N, Gaulton A, et al. Pharos: Collating protein information to shed light on the druggable genome. Nucleic acids research. 2017;45(D1):D995-d1002. Epub 2016/12/03. doi: 10.1093/nar/gkw1072. PubMed PMID: 27903890; PubMed Central PMCID: PMCPMC5210555.

7. Kinoshita J, Clark T. Alzforum. Methods in molecular biology (Clifton, NJ). 2007;401:365-81. Epub 2008/03/28. doi: 10.1007/978-1-59745-520-6_19. PubMed PMID: 18368375.