

Principal Investigator

First Name: David

Last Name: Ennist

Degree: PhD, MBA

Primary Affiliation: Origent Data Sciences, Inc

E-mail: dennist@origent.com

Phone number: 301-466-7693

Address: Origent data Sciences, Inc., 1875 Connecticut Ave NW, 10th floor, Washington, DC 20009

Origent data Sciences, Inc., 1875 Connecticut Ave NW, 10th floor

City: Washington

State or Province: DC

Zip or Postal Code: 20009

Country: United States

SCOPUS ID: 6602119295

General Information

Key Personnel (in addition to PI):

First Name: David

Last name: Ennist

Degree: PhD

Primary Affiliation: Origent Data Sciences, Inc.

SCOPUS ID: 6602119295

First Name: Jonavelle

Last name: Cuerdo

Degree:

Primary Affiliation: Origent Data Sciences

SCOPUS ID:

Are external grants or funds being used to support this research?: External grants or funds are being used to support this research.

Project Funding Source: National Institute on Aging

How did you learn about the YODA Project?: Other

Conflict of Interest

https://yoda.yale.edu/system/files/2020-04-23_yoda_project_coi_form_for_data_requestors_2019.pdf.pdf

https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_origent_cuerdo.pdf

Certification

Certification: All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.

Data Use Agreement Training: As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training

1. [NCT00679627 - GALALZ3005 - A Randomized, Double-Blind, Placebo-controlled Trial of Long-term \(2-year\) Treatment of Galantamine in Mild to Moderately-Severe Alzheimer's Disease](#)

2. [NCT00575055 - ELN115727-302 - A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled,](#)

- [Parallel-Group, Efficacy and Safety Trial of Bapineuzumab \(AAB-001, ELN115727\) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Carriers](#)
3. [NCT00574132 - ELN115727-301 - A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy and Safety Trial of Bapineuzumab \(AAB-001, ELN115727\) In Patients With Mild to Moderate Alzheimer's Disease Who Are Apolipoprotein E4 Non- Carriers](#)

What type of data are you looking for?: Individual Participant-Level Data, which includes Full CSR and all supporting documentation

Research Proposal

Project Title

Machine Learning Applications for Improving the Efficiency of Clinical Trials in Alzheimer's Disease

Narrative Summary:

This proposal aims to develop machine-learning predictive models and applications that increase the efficiency of drug development clinical trials for Alzheimer's disease. We propose to improve and validate our prototype ADAS-Cog 11 model, develop and validate models for additional commonly used AD trial endpoints, make the model outputs available in real-time to clinical trialists through an application programming interface, fully develop the AD clinical trial applications and prepare the infrastructure and documentation needed to support regulatory submissions. These models and applications will vastly increase the speed and efficiency of drug development for Alzheimer's disease.

Scientific Abstract:

Background: Alzheimer's disease (AD) is characterized by a degree of heterogeneous disease progression that has been cited by numerous authors as a contributing factor in the dismal record of drug development clinical trials for AD. This complex disease is influenced by an interplay between several genetic and environmental factors that have been difficult to capture using traditional modeling techniques. It is clear that drug development for AD would benefit from a framework that stratifies AD patients by predicted disease progression, presenting the possibility of enriching clinical trials with homogeneous participants and using the predictions as covariates for improved randomization and covariate adjustment. This grant application builds on our previous work that resulted in the creation of robust, commercializable machine learning-based clinical trial applications currently being used in drug trials for amyotrophic lateral sclerosis (ALS). We seek to develop similar applications to increase the efficiency of drug development clinical trials for AD.

Objective: To increase the efficiency of clinical trials in AD.

Study design: To refine our current machine learning ADAS-cog model, develop an MMSE model, validate the models with external datasets, run simulations including virtual controls and power analyses.

Participants: Critical Path Institute CAMD and ADNI datasets.

Main outcome measures: ADAS-cog and MMSE

Statistical analysis: Models evaluated using root-mean square deviation, R^2 , bias, receiver operating characteristics curve, discrimination & calibration.

Brief Project Background and Statement of Project Significance:

Responses to 7/7 email:

1. Thank you, we are aware of the DUA. The three Alzheimer's disease trial sets requested will be used to validate models already developed using ADNI and Critical Path Institute datasets. Our intent is to present the results at scientific meetings and publish our results in Peer-reviewed journals.
2. The models have already been developed and the three trials will be used for external validation. We will perform analyses using the training set via 10 fold cross validation, generating root-mean square deviation to assess accuracy and we will examine mean prediction error for bias analysis. From our extensive ALS studies, we have found empirically that an RMSD within 15% of the scale being predicted gives us results useful for the applications we've developed - including virtual controls, enrichment, stratification, randomization, covariate adjustment and

subgroup analysis.

Our licenses with CPI and ADNI prevent us from disseminating the data, so we will not upload the data sets into the YODA environment.

Predictors include ADAS total score, sex, baseline word recall, visit study day, word recognition, orientation, object naming, commands, test instruction recall, MMSE score, age, ideational praxis, several labs, spontaneous speech word finding, spoken language, weight, pulse, height, blood pressure.

In our successful phase 2 SBIR grant, we used ALS as a model disease to develop our ALS product based on machine learning ALS disease progression models, which generate the necessary data for use as virtual controls and for stratifying patients for either trial enrichment or randomization, or for subsequent use as covariates in the statistical analysis of a trial. We successfully commercialized our ALS product prototype into a robust, scalable, market-ready ALS product we call ForecastOne® which includes a versatile Application Programming Interface (API) that can be integrated with the electronic data capture (EDC) systems of clinical trial to both return predictions in real-time and to store the predictions in the trial database. Our ForecastOne® API is currently in use in a clinical trial being conducted by one of our pharma clients at the Massachusetts General Hospital.

In addition to ALS, we have developed a prototype model for a common AD trial endpoint, the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-cog) and showed the potential of the model to lower sample size and boost the power of AD clinical trials. This ADAS-cog prototype model serves as the completed proof-of-concept starting point for this Direct-to-Phase 2 SBIR grant application. The objectives of this grant are to improve the prototype ADAS cog model, develop a Mini-Mental State Examination (MMSE, another common AD trial endpoint) model, validate the models with external datasets, make them accessible to clinical trialists through our API, and create several applications for AD drug development using the models, including enrichment, randomization, covariate adjustment and virtual controls. The applications will be made ready to use in AD clinical trials.

Specific Aims of the Project:

AIM 1: Improve & validate our prototype ADAS-Cog 11 model and develop & validate ADAS-Cog 13 & MMSE models.

Challenge: Develop a robust, scalable AD product that makes predictions for commonly used AD outcomes available to clinical trialists directly through their electronic data capture systems, in real-time without multiple data entry.

Aim 2: Develop AD applications, including stratification, randomization, covariate adjustment & prognostic matching.

Challenge: Provide evidence through simulations that the use of the predictions will increase the efficiency of AD clinical trials.

Aim 3: Prepare Open API infrastructure and documentation necessary for inclusion of predictive strategies in trial regulatory submissions for Alzheimer's disease.

Challenge: Create a robust framework compliant with regulatory needs for the use of predictions in the AD drug candidate submissions of our clients.

What is the purpose of the analysis being proposed? Please select all that apply.

Develop or refine statistical methods

Research on clinical trial methods

Research on clinical prediction or risk prediction

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:

Critical Path Institute CAMD dataset and ADNI datasets. Selected participants must have longitudinal a record, baseline ADAS-cog & MMSE, date of onset and date of diagnosis.

Main Outcome Measure and how it will be categorized/defined for your study:

Main outcomes to be analyzed are ADAS-Cog 11 and MMSE, as recorded in the CAMD and ADNI participant records.

Main Predictor/Independent Variable and how it will be categorized/defined for your study:

All discovery will be performed using the CAMD and ADNI datasets that we have already obtained. Once we have discovered all the important predictors, built the models and validated them using internal 10 fold cross validation, we will use the data sets being requested here to validate the models and applications. We will first go through a variable reduction effort using the random forest algorithm. Preliminary studies indicate that baseline ADAS-Cog 11 is the best predictor for ADAS-Cog 11; likewise, baseline MMSE is likely to be the best predictor for MMSE.

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:

Demographics, labs, ApoE status, vitals, time since symptom onset, time since diagnosis. In addition, we hope to discover additional useful features with the CAMD and ADNI datasets.

Statistical Analysis Plan:

We will characterize the models using 10 fold cross validation using 10% set aside of the training dataset (merged CAMD and ADNI datasets).

Regression models will be characterized using RMSD, R² and bias analysis.

Time to event models will be characterized using ROC curves, discrimination and calibration.

The clinical trial datasets requested here will be used for external validation using the same metrics. They will not be used for model training.

Software Used:

RStudio

Project Timeline:

The NIA SBIR grant is expected to begin in 1/21 and continue for 2 years. We request access to the requested datasets as soon as possible to align with our structures prior to funding. Anticipate publication submitted 6/22 to Neurology. Results reported to YODA 3/23.

The project timelines are outlined below:

Table 4. Timelines (project quarter) 1 2 3 4 5 6 7 8

AIM 1: ADAS-Cog 11 model development X X X

AIM 1: ADAS-Cog 13 model development X X X

AIM 1: MMSE model development X X X

AIM 2: Develop AD prognostic match application X X X

AIM 2: Prognostic match simulations X X X

AIM 2: Stratification simulations X X X X X

AIM 2: Randomization simulations X X X X

AIM 2: Covariate adjustment simulations X X X X

AIM 3: Model publication to API X X X X

AIM 3: Quality control X X X X X X X

AIM 3: Documentation X X X X X X X

AIM 3: GDPR & 21 CFR part 11 Compliance X X

Dissemination Plan:

We expect that the applications we develop will be useful for drug development. The datasets being requested here will be used as external datasets solely to validate previously trained models. The datasets will not be used in model building, only for validation, and they will thus not be included in the anticipated products. The target audiences of the products are academic, pharma and biotech Sponsors of Alzheimer's disease clinical trials. Our team includes Dr. David Bennett, Director of the Alzheimer's Disease Center of Rush University Medical Center, Chicago - we plan to publish the results in Neurology and present at AAIC 2022 and 2023.

Bibliography:

Our previous work has been in amyotrophic lateral sclerosis (ALS). We intend to apply many of the same methods to the development of AD models. Here are publications of our ALS studies:

1. Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, Li G, Fang L, Mackey L, Hardiman O, Cudkovicz M, Sherman A, Ertaylan G, Grosse-Wentrup M, Hothorn T, van Ligtenberg J, Macke JH, Meyer T, Schölkopf B, Tran L, Vaughan R, Stolovitzky G, Leitner ML. Crowdsourced analysis of clinical trial data to predict amyotrophic

- lateral sclerosis progression. *Nat Biotechnol.* 2015; 33(1): 51-57. doi: 10.1038/nbt.3051
2. Zach N, Ennist DL, Taylor AA, Alon H, Sherman A, Küffner R, Walker J, Sinani E, Katsovskiy I, Cudkowicz M, Leitner ML. Being PRO-ACTive: What can a Clinical Trial Database Reveal About ALS? *Neurotherapeutics.* 2015;12(2):417-23. doi: 10.1007/s13311-015-0336-z
 3. Taylor AA, Fournier C, Polak M, Wang L, Zach N, Keymer M, Glass JD, Ennist DL. Predicting Disease Progression in Amyotrophic Lateral Sclerosis. *Ann Clin Transl Neurol.* 2016; 3:866-875. doi: 10.1002/acn3.348.
 4. Jahandideh S, Taylor AA, Beaulieu D, Keymer M, Meng L, Bian A, Atassi N, Andrews J, Ennist DL. Longitudinal modeling to predict vital capacity in amyotrophic lateral sclerosis, *Amyotroph Lateral Scler Frontotemporal Degener.* 2018; 19:294-302. doi: 10.1080/21678421.2017.1418003
 5. Berry JD, Taylor AA, Beaulieu D, Meng L, Bian A, Andrews J, Keymer M, Ennist DL, Ravina B. Improved stratification of ALS clinical trials using predicted survival. *Ann Clin Transl Neurol.* 2018; 5(4): 474–485. doi: 10.1002/acn3.550
 6. Nicholson K, Chan J, Macklin EA, Levine-Weinberg M, Breen C, Bakshi R, Grasso DL, Wills A-M, Jahandideh S, Taylor AA, Beaulieu D, Ennist DL, Andronesi O, Ratai E-M, Schwarzschild MA, Cudkowicz M, Paganoni S. Pilot trial of inosine to elevate urate levels in amyotrophic lateral sclerosis. *Ann Clin Transl Neurol.* published online Nov 12, 2018. doi: 10.1002/acn3.6871
 7. Schoenfeld DA, Finkelstein DM, Macklin E, Zach N, Ennist DL, Taylor AA, Atassi N, The Pooled Resource Open-Access ALS Clinical Trials Consortium. Design and analysis of a clinical trial using previous trials as historical control. *Clinical Trials.* First Published online 1 Jul 2019. doi: 10.1177/1740774519858914
 8. Ennist DL, Beaulieu D, Taylor AA, Pierce D, Cuerdo J, Keymer M. (2020). Improving Clinical Trial Efficiency with Machine Learning Models of Disease Progression, In RA Smith, B Kaspar & C Svendsen, (Eds.), *Neurotherapeutics in the Era of Translational Medicine.* Elsevier (in press).

Abstracts:

1. Beaulieu D. Data Delivery: From Paper to Pipeline Using R. Conference on Statistical Practice, Pipeline and Parallel Computing Using R, February 21, 2020, Sacramento, CA. link
2. Brooks BR, Piro E, Schactman M, Beaulieu D, Taylor AA, Keymer M, Agnese W, Perdrizet J, Apple S, Ennist DL. Evidence for Generalizability of Edaravone Efficacy Using a Novel Machine-Learning Risk-Based Analysis Tool. 30th International Symposium on ALS/MND; December 4-6, 2019; Perth, Australia. Theme 1 Epidemiology and informatics, Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 20:sup1, page 108, abstract EPI-13; doi: 10.1080/21678421.2019.1646989
3. Beaulieu D, Cuerdo J, Taylor AA, VanMeter S, Zhao E, Keymer M, Ennist DL. Estimate of an Acthar® Gel Treatment Effect in ALS Patients using Virtual Controls. 30th International Symposium on ALS/MND; December 4-6, 2019; Perth, Australia. Theme 9 Clinical trials and trial design, Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 20:sup1, page 278, abstract CLT-27; doi: 10.1080/21678421.2019.1646997
4. Taylor AA, Beaulieu D, Cuerdo J, Pierce D, Conklin A, Keymer M, Ennist DL. Detectable Effect Cluster Analysis: A Novel Machine Learning Based Clinical Trial Subgroup Analysis Tool. 30th International Symposium on ALS/MND; December 4-6, 2019; Perth, Australia. Theme 9 Clinical trials and trial design, Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration, 20:sup1, page 265, abstract CLT-06; doi: 10.1080/21678421.2019.1646997
5. Brooks BR, Piro E, Schactman M, Beaulieu D, Taylor AA, Keymer M, Agnese W, Perdrizet J, Apple S, Ennist DL. Evidence for Generalizability of Edaravone Efficacy Using a Novel Machine-Learning Risk-Based Analysis Tool. Muscle Study Group Annual Meeting, Snowbird, UT, September 19-22, 2019. (presented as a talk in Sponsor Update Session). *Muscle & Nerve*, 60:supS2, November 2019, page S11, doi: 10.1002/mus.26666
6. Beaulieu D, Taylor AA, Conklin A, Cuerdo J, Pierce D, Keymer M, Ennist DL. Detectable Effect Cluster Analysis: A Novel Machine-Learning Based Clinical Trial Subgroup Analysis Tool. 18th Annual NEALS Meeting, Tampa, FL, USA, October 2019. *F1000Research* 2019, 8:1883 (POSTRET; doi: 10.7490/f1000research.1117619.1
7. Brooks BR, Piro E, Schactman M, Beaulieu D, Taylor AA, Keymer M, Agnese W, Perdrizet J, Apple S, Ennist DL. Evidence for Generalizability of Edaravone Efficacy Using a Novel Machine-Learning Risk-Based Analysis Tool. 18th Annual NEALS Meeting, Tampa, FL, USA, October 2-4, 2019; abstract 76.
8. Taylor AA, Beaulieu D, Pierce D, Conklin A, Cuerdo J, Keymer M, Ennist DL. Rapid deployment of a Machine Learning-based derived biomarker using publicly available data sources for covariate adjusted descriptive modeling. Symposium on Data Science and Statistics, Machine Learning, May 31, 2019, Bellevue, WA. link
9. Beaulieu D, Taylor AA, Cuerdo J, Conklin A, Keymer M, Ennist DL. Increasing ALS Clinical Trial Efficiency using Machine Learning Models. ENCALS Meeting 2019, Tours, France; May 15-17, 2019; abstract 1918.
10. Beaulieu D, Taylor AA, Cuerdo J, Conklin A, Keymer M, Ennist DL. Increasing ALS Clinical Trial Efficiency using Machine Learning Models. 71st Annual Meeting of the American Academy of Neurology, Neuromuscular and Clinical Neurophysiology Topic, Philadelphia, PA; May 5-10, 2019; abstract P1.4-021.
11. Beaulieu D, Taylor AA, Cuerdo J, Conklin A, Keymer M, Ennist DL. Increasing ALS Clinical Trial Efficiency

- using Machine Learning Models. 2019 MDA Clinical and Scientific Conference, Orlando, FL; April 13-17, 2019.
12. Beaulieu D, Cuerdo J, Taylor AA, Bian A, Meng L, Wolff AA, Ennist DL. Validation of a Suite of Machine Learning Models using the Longitudinal VITALITY-ALS Data Set. 29th International Symposium on ALS/MND; Glasgow, Scotland; Dec 5, 2018. F1000Research 2018, 7:1944 (poster; doi: 10.7490/f1000research.1116358.1)
 13. Beaulieu D, Taylor AA, Cuerdo J, Conklin A, Keymer M, Ennist DL. Machine Learning Applications for Increasing the Efficiency of ALS Clinical Trials. 17th Annual NEALS Meeting, Tampa, FL, USA, October 3, 2018. F1000Research 2018, 7:1782 (poster; doi: 10.7490/f1000research.1116286.1)
 14. Beaulieu D, Taylor AA, Conklin A, Cuerdo J, Keymer M, Ennist DL. Increasing Study Power using a Machine Learning Approach. Joint Statistical Meeting (JSM) of the American Statistical Association (ASA) in Vancouver, British Columbia, Canada, July 30, 2018. F1000Research 2018, 7:1785 (poster; doi: 10.7490/f1000research.1116287.1)
 15. Beaulieu D, Taylor AA, Conklin A, Ennist DL. Machine Learning Tools for Improving the Efficiency of Drug Development Clinical Trials in ALS. ENCALS Meeting 2018, Oxford, UK; June 20-22, 2018; abstract D27
 16. Ennist DL, Beaulieu D, Jahandideh S, Taylor AA. Machine Learning Tools for Improving the Efficiency of Drug Development Clinical Trials in ALS. AAN 2018 Meeting, Los Angeles, CA; April 21-27, 2018
 17. Taylor AA, Beaulieu D, Jahandideh S, Meng L, Bian A, Andrews J, Ennist DL. Validation of Predictive ALS Machine Learning Models with a Contemporary, External Dataset and Application to Trial Simulations. 28th International Symposium on ALS/MND; 2017 December 9; Boston, MA. F1000Research 2018, 7:322 (poster; doi: 10.7490/f1000research.1115313.1) (also presented at NEALS 2017)
 18. Jahandideh S, Ennist DL. Machine Learning Models for the Assessment of Potential ALS Biomarkers. 28th International Symposium on ALS/MND; 2017 December 9; Boston, MA. F1000Research 2018, 7:322 (poster; doi: 10.7490/f1000research.1115312.1)
 19. Fournier CN, Taylor AA, Ennist DL. Enriched clinical trial cohorts improve study power. 16th Annual NEALS Meeting, Tampa, FL, USA, October 4, 2017. F1000Research 2017, 6:1868 (poster; doi: 10.7490/f1000research.1114995.1)
 20. Taylor AA, Beaulieu D, Jahandideh S, Meng L, Bin A, Andrews J, Ennist DL. Validation of predictive ALS machine learning models with a contemporary, external dataset and application to trial simulations. 16th Annual NEALS Meeting, Tampa, FL, USA, October 4, 2017. F1000Research 2017, 6:1867 (poster; doi: 10.7490/f1000research.1114992.1) (also presented at ALS/MND 2017)
 21. Taylor AA, Jahandideh S, Beaulieu D, Keymer M, Ennist DL. Machine learning models for the clinical development of gene and cell therapies. 20th Annual Meeting of The American Society for Cell and Gene Therapy, May 10, 2017. F1000Research 2017, 6:1866 (poster; doi: 10.7490/f1000research.1114991.1)
 22. Bedlack RS, Ennist DL, Taylor A and Keymer M. ALS resistance is regional and not explained by demographics, medications or labs. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2016, 5:2917 (poster; doi: 10.7490/f1000research.1113569.1)
 23. Machine learning model for the prediction of slow vital capacity. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2017, 6:8 (poster) (doi: 10.7490/f1000research.1113590.1)
 24. Miller RG, Ennist DL, Jenkins L, Thompson J, Fritz M, Moore DH. Novel trial design in a clinical study of Diaphragm Pacing (DPS) for ALS. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2017, 6:9 (poster) (doi: 10.7490/f1000research.1113591.1)
 25. Taylor AA, Fournier C, Polak M, Wang L, Zach N, Shepperson J, Reichert J, Keymer M, Glass JD, Ennist DL. Predicting disease progression for ALS clinic patients. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2017, 6:4 (poster; doi: 10.7490/f1000research.1113586.1)
 26. Taylor AA, Jahandideh S, Beaulieu D, Keymer M, Ennist DL. In silico stratification of ALS patients using machine learning algorithms. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2017, 6:3 (poster; doi: 10.7490/f1000research.1113585.1)
 27. Schoenfeld D, Küffner R, Macklin E, Ennist DL, Moore DH, Zach N, Atassi N. The proper use of historical controls in ALS trials. 27th International Symposium on ALS/MND; 2016 December 9; Dublin, Ireland. F1000Research 2016, 5:2904 (poster; doi: 10.7490/f1000research.1113562.1)
 28. Taylor AA, Miller R, Onders R, Ennist DL. Analysis of function and survival in ALS patients with diaphragm pacing using virtual controls. 26th International Symposium on ALS/MND; 2015 December 11; Orlando, FL, USA. F1000 Research 2016. 5:120 (poster; doi: 10.7490/f1000research.1111282.1)